# Private Access Fees and Congestion: Is There a Role for Government After All?

Stephen Salant* Nathan Seegert†‡

July 2014

**Abstract**

We reconsider an important debate between Pigou (1920) and Knight (1924) nominally about congestible roads. Contrary to Knight's contention, allowing independent players to set tolls on congestible roads does not necessarily induce an efficient allocation of motorists. Toll-setting does result in efficient allocation in the limit of a large economy even in the absence of any uncongestible road. But in the more realistic circumstance of a finite economy, toll-setting—unlike Pigouvian taxes—will not in general achieve efficiency. We use these results to demonstrate a role for government in producing an uncongestible option that provides the necessary "competitive conditions" for toll-setting to reach the efficient outcome. These results are important for other allocation problems involving congestion. They even apply to the allocation of researchers across different contests where prize setters offer each winner a monetary prize and must compete with other prize setters and with jobs offering riskless compensation.

**Keywords:** Congestion externality
**JEL Classification:** H21, H23.

*Department of Economics, University of Michigan

†nathan.seegert@business.utah.edu. Department of Finance, University of Utah

‡We would like to thank Natalia Lazzati and Molly Macauley for comments on a previous draft.

# 1 Introduction

Consider a market where $N$ identical buyers each purchase one unit of a good from one of $n$ heterogeneous producers. Producers set prices and sell imperfect substitutes. A seller can steal a fraction of his rivals' customers by marginally reducing his price. Buyers impose externalities on each other in the sense that, at fixed prices, a decision by one buyer can harm other buyers. Given the externalities and the price-setting, the sufficient conditions of the First Welfare Theorem are clearly violated, and, from the perspective of the Arrow-Debreu model, efficiency of the market allocation seems unlikely.

Knight (1924), one of the intellectual founders of the Chicago School, provided an example of an economy with both price-setting and a particular class of externalities but where the decentralized equilibrium is nonetheless efficient. Knight considered the allocation of motorists (buyers) between a free, uncongestible road and a faster road subject to a congestion externality. Pigou (1920) had pointed out that if each motorist simultaneously chooses his own route to work, the resulting Nash equilibrium allocation of motorists fails to minimize the total wages lost commuting since a motorist choosing the congestible road disregards the impact of his choice on the commute time of fellow motorists. Pigou had suggested that efficiency could be restored through a government tax at the "Pigouvian level" on motorists using the congestible road. In response, Knight pointed out that government intervention is unnecessary: if ownership of the congestible road were assigned to a self-interested private agent, the owner would charge motorists the revenue-maximizing toll for access and an efficient allocation would result without any government intervention.

Knight appeared to win the debate. Pigou dropped the congestible roads example from subsequent editions of his textbook, *The Economics of Welfare*.[1] In our view, Pigou's apparent capitulation in this debate was far too hasty.

Knight's idea has been applied in countless other contexts. If fishermen can choose whether to fish on one of several congestible lakes or to take a job at a low fixed wage, for example, the fishermen will allocate themselves among the lakes and wage job in a way that equalizes earnings from each activity. But this allocation fails to maximize their aggregate earnings since no one takes account of the negative impact his own presence has on the productivity of the others fishing on the same lake. Instead of a lake-specific government tax to restore efficiency, however, disciples of Knight recommend empowering one individual per lake to set a fee that must be paid to gain access to the lake since the revenue-maximizing fee on each lake would supposedly be set equal to the lake-specific Pigouvian tax.

[1] Cheung (1973) (footnote 2) interprets Pigou's withdrawal of the highway example from his opus as an attempt to avoid further criticism by Knight.

Or, to take a thoroughly modern example, suppose there are a given number of biomedical researchers who can be deployed to discover the cures for a set of diseases or alternatively can work for NIH on some other socially valuable project. Discovering a cure for each disease has a specified social value as does working for NIH. This is a congestion problem since when any researcher seeks the cure for a given disease he reduces the chance that another person will be the first to discover that cure, an effect that the planner would take into account but that a private agent would ignore when deciding which activity to pursue. A planner could work out how many of the researchers should be deployed in each activity to maximize expected social welfare and the government could set "Pigouvian prizes" to achieve this efficient allocation. However, disciples of Knight would argue that the government is not needed; private agents would find it in their own interest to offer Pigouvian prizes to the first discoverer of cures in exchange for the property right to produce the resulting drug. Will competing prize setters in theory achieve the efficient allocation?[2]

In this paper, we re-examine Knight's claim that competition—among highway toll setters, among gate keepers charging for access to lakes, or among contest prize setters—will eliminate allocative inefficiency, making government intervention unnecessary. In the body of the paper, we will confine our attention to Knight's original context—highway congestion. We will return to these other applications in the concluding section and the appendix.

In the context of highway congestion, we show that—despite the price-setting and the negative externalities—toll-setting *does* result in efficient allocation in the limit of a large economy even in the absence of any uncongestible road. Thus, when suitably extended, Knight's example is an interesting complement to Edgeworth's (1881) result about core convergence in replica economies. But in the more realistic circumstance of a finite economy, toll-setting—unlike Pigouvian taxes—will not in general achieve efficiency. This is in contrast to Demsetz's (2010) suggestions that toll-setting is efficient if there are two roads where "one road is subject to considerable congestion because it is narrow; the other road is wide and escapes much of this congestion ..." (Demsetz, 2010, p. 4).

For efficiency in a finite economy, one needs an uncongestible road—and not just any uncongestible road. The road must be of "sufficient speed" that the toll setters will not misbehave. Our analysis clarifies the positive and normative consequences of toll-setting as the speed of the uncongestible road is varied exogenously.

What can be done if no uncongestible road of sufficient speed is available? Pigouvian taxes are, of course, an option. But even if the government lacks the detailed information necessary to set a highway-specific tax on every road, it will likely have the information necessary to provide an

[2]Prizes are becoming an important mechanism in many areas. See Brennan et al. 2011 for a partial list.

uncongestible alternative speedy enough to induce efficient toll-setting. Thus, the government and the toll setters would together induce motorists to allocate themselves efficiently.

This result suggests that the government can sometimes improve allocative efficiency by supplying consumers with an alternative to what the private sector provides. If the congestion in the busy New York–Washington corridor was resolved by allowing independent toll-setting on alternative congestible routes, allocative efficiency could be improved if the government provided a train running with sufficient frequency that motorists considered it a viable alternative to driving on congested tollroads and then parking in congested lots. The express package delivery services (FedEx, UPS, DHL, etc.) may function more efficiently if the U.S. postal service provided consumers with a viable alternative, and so forth.

We proceed as follows. In the next section, we introduce notation and set up the model. In section 3, we exposit Knight's original contribution with one toll-setter and an uncongestible road attractive enough to be utilized and then generalize it to $n$ congestible toll roads and a utilized outside option. In section 4, we clarify that with no outside option (or, equivalently, with a sufficiently unattractive outside option), the allocation need not be efficient in a finite economy, although efficiency would be restored if the economy were replicated sufficiently. Section 5 contains a positive and normative analysis of duopoly toll setting as the attractiveness of the outside option is varied exogenously. In section 6, we use our findings to suggest a new role the government can play in a mixed economy: supplying an uncongestible alternative sufficiently attractive to provide either actual or potential competition to alter the behavior of private toll setters. Section 7 discusses comparative statistic and section 8 concludes by pointing out the implications of our analysis for a quite different congestion problem.

## 2 Related Literature

Developing mechanisms to provide roads and solve the problem of congestion has its beginning with the French engineers of the 1840s-1850s (Dupuit 1844) and early 20th century economists Pigou (1920) and Knight (1924). Following Pigou's argument for government taxes and Knight's 1924 rebuttal in favor of private toll setters, a large literature has developed on both sides (Lindsey 2006). Early work by Buchanan (1956) and Mills (1981) identified that Knight's argument depended upon competition among private road suppliers because in the absence of competition private road suppliers would have some degree of monopoly power and limit traffic excessively (Edelson 1971).

Seminal papers by Vickrey (1963) and Walters (1961) formalized Pigou's initial argument that roads are goods that are misallocated in the absence of a market. Buchanan (1965) re-imagines this argument by framing roads as club goods, an intellectually attractive approach which many

others followed Ellickson, Grodal, Scotchmer and Zame (1999).[3] Following Knight's critique work by Rothbard (1978), Block (1979, 1980, 1996), and Cadin and Block (1997) continued to describe the benefits private roads would have on efficient capacity and operations. Fielding and Klein (1993) re-imagines the private roads argument suggesting franchising as a practical mechanism. These arguments are continued by Roth (2006) Klein and Majewski (2006), Button (2006), and Foldvary (2006). The support of privatization of roads also stems from work following Coase (1960) (Winston and Yan 2011).[4]

Later work focuses on second-best approaches to congestion including when there exists an unpriced alternative (Verhoef, Nijkamp and Rietveld 1997, Small and Yan 2001), complementary goods such as parking (Arnott, De Palma and Lindsey 1991) and other transportation modes (Parry and Small 2009).[5] A large literature also exists on dynamic congestion models, initially developed by Henderson (1974) and Edelson and Hilderbrand (1975), and bottleneck models, developed by Vickrey (1969) and extended to include additional features such as heterogeneous commuters and time varying tolls (Arnott, De Palma and Lindsey 1988, 1990, 1992, 1993, 1994, and 1998).

## 3 Model Setup

Pigou and Knight both considered $N$ motorists trying to get from the same starting point to the same destination using one of several roads. There are $N$ motorists, each of whom earns the same wage. The time cost of delay in each motorist's commute is measured in dollars of forgone wages. Pigou and Knight classified each road the motorists might take as either uncongestible or congestible. An uncongestible road provides a constant commute time (denoted $\$c$) regardless of the number of drivers on the road. The commute time on a congestible road depends on the number of drivers using the same road. Denote the cost per motorist to commute on road $i$ (again denominated in dollars) as $A_i(x_i)$ where $x_i$ is the number of drivers on road $i$.[6] We assume below (1) that $A_i(\cdot)$ is differentiable, (2) that it is strictly increasing, and (3) that $x_i A_i(x_i)$ is strictly convex for $i = 1, \ldots, n$. Thus, the more drivers drive on road $i$ the higher cost each of them pays

[3]The competitive equilibria of clubs is also discussed by Berglas (1976), Berglas (1981), Berglas and Pines (1981), Boadway (1980), Berglas, Helpman and Pines (1982), Hillman and Swan (1983), Sandler and Tschirhart (1980), Scotchmer (1994), and Scotchmer and Wooders (1987). Seegert (2011) discusses the equilibria in clubs in the context of cities.

[4]There is a large literature following Coase (1960) that considers the problem of congestion as one of ill defined property rights. Solutions include transferable discharge permits (Crocker 1966, Dales 1968) accounting for concentration in the permit and product market (Malueg 1990) and transaction costs (Stavins 1995).

[5]There exists a large group of empirical research on traffic congestion, most recently Anderson (2013) and Hughes and Kaffine (2013).

[6]Throughout we treat the number of drivers on road $i$ (denoted $x_1$) as a real number rather than an integer.

in lost wages. Throughout the paper we assume that if there is an uncongestible road its commute time is between the commute time on any congestible road when no motorists are on the road, $c > A_i(0)$, and when all motorists are on the road $c < A_i(N)$.[7]

Pigou and Knight considered the allocation of the $N$ motorists to the available roads that minimized total man-hours lost to the commute and compared it to the decentralized solution that would occur if each motorist chose her own road. Formally, Knight envisioned a two-stage game where the toll was set at the first stage and the $N$ motorists, after observing it, chose their routes in the second stage. Let $\theta_1$ denote the toll. In the second stage motorists allocate themselves such that the total costs of commuting are equal, $A_i(x_i) + \theta_i = c$ or $A_i(x_i) + \theta_i = A_j(x_j) + \theta_j$, depending on whether the uncongestible road exists at a sufficient speed. We assume that it is never advantageous for the toll setter to drive *every* motorist away from his road or attract every motorist away from the uncongestible road.[8]

We define the socially optimal solution (denoted $x_i^s$) as the allocation of motorists to the $n$ roadways that minimizes the total cost of commuting, $\sum_{i=1}^{n} x_i A_i(x_i) + (N - \sum_{j=i}^{n} x_i)c$. Given our assumptions, the minimand in the planning problem is strictly convex so the solution $(\{x_i^s\}_{i=1}^n)$ of the $n$ first-order conditions is unique. Equivalently, the planner maximizes the time saved over sending everyone on the uncongestible, slow road: $cN - \{\sum A_i(x_i)x_i + (N - \sum x_i)c\}$. The planner maximizes the time saved by minimizing the commute time (the expression in braces). Alternatively, by using road $i$ instead of the uncongestible road, $x_i$ motorists save $R_i(x_i) = x_i(c - A_i(x_i))$. The planner maximizes the aggregate time saved: $\sum_i R_i(x_i)$.

The solution to the planner's problem is to set the marginal cost equal across all roads, $A_i(x_i^s) + x_i^s A_i'(x_i^s) = A_j(x_j^s) + x_j^s A_j'(x_j^s) = c$. Pigou and Knight compared this efficient benchmark to the Nash equilibrium that would occur if each motorist simultaneously minimized the cost of commuting to work, taking into account any taxes and tolls on the different routes. As we will show in the following sections the socially optimal allocation can be achieved using a "Pigouvian" tax $\theta_1^P = x_1^s A_1'(x_1^s)$ or in some cases by allowing revenue-maximizing toll setters to set a toll.

## 4 Review of the Knight-Pigou Controversy

Pigou pointed out that if there were two roads, one congestible ($n = 1$) and one uncongestible, the free movement of drivers across the two roads would lead to the commute times being equal,

[7] We discuss the role of the uncongestible road in section 6 by exogenously varying its commute time.

[8] Intuitively, driving every motorist away generates no toll revenue while charging so little that every motorist strictly prefers one's highway is dominated by charging slightly more than what would make motorists indifferent between one's own roadway and the next best alternative.

$A_1(x_1) = c$, despite the fact that the socially optimal allocation prescribes $A_1(x_1^s) < c$. This market inefficiency is due to an externality each driver on the congestible road imposes on the other drivers on the congestible road by slowing down traffic. To restore efficiency, Pigou proposed that the government set a toll per motorist of $\theta_1^P = x_1^s A_1'(x_1^s)$ for each motorist using the congestible road. Faced with this "Pigouvian" tax, drivers allocate themselves across the two roads such that marginal benefits on each road are equalized: $A_1(x_1) + x_1^s A_1'(x_1^s) = c$, which insures efficiency.

Knight replied that government intervention was unnecessary to achieve efficiency. If the road was privately owned and its owner charged an entry fee that maximized his revenue, Knight showed that the toll setter would inadvertently induce the efficient allocation of motorists between the two roads. To demonstrate Knight's claim consider the objective of the toll setter maximizing revenue from the road:

$$\pi_1(\theta) = \theta_1 x_1(\theta_1) = x_1(\theta_1)(c - A_1(x_1(\theta_1)) \text{ with respect to } \theta_1. \tag{1}$$

Inadvertently the toll setter maximizes $R(x_1)$, which is the savings in commute time from using the congestible road and the social planner's objective function, as shown in section 3. Hence, the toll setter's objective function is a strictly convex function which achieves its global maximum at $x_1^s$. One intuition for Knight's result is that for any $\theta_1$ chosen by the toll setter, the payoff of the $N$ motorists does not change: collectively, the motorists always lose \$Nc (in wages and tolls) and the other toll setters receive the same payoffs. So when the toll setter varies his decision variable to maximize his own payoff, he inadvertently maximizes the sum of everyone's payoff.[9]

Suppose instead there were $n$ roads congestible to different degrees and one uncongestible road. Self-interested competitive toll setters will again inadvertently induce self-interested motorists to allocate themselves in a way that minimizes the total time lost commuting. Once again, private toll setters would set exactly the same tolls as the government tax authority.

To verify this claim, assume that every toll setter chooses the Pigouvian toll (denoted $\theta_j^P$) where $\theta_j^P = x_j^s A_j'(x_j^s)$ for $j = 1, \ldots, n$. Suppose one of the $n$ players (player $i$) considers a unilateral deviation. If he lowers his toll, motorists are attracted to his roadway until the increased congestion makes it no longer more attractive than the uncongestible, toll-free road; if he raises his toll, motorists leave his roadway until its reduced congestion makes it just as attractive as the uncongestible, toll-free road. When he varies his toll, motorists reallocate themselves between his road and the uncongestible road. But as long as any motorist remains on the uncongestible road, changes in his toll affect neither (1) the number of motorists using each of the other $n-1$ congestible

[9]The maximum point found in this section assuming a "sufficiently fast" uncongestible road is the global maximum, which we show in Appendix A.

roads nor (2) the revenue each of the $n-1$ other toll setters collects.

As long as any motorist remains on the uncongestible road $A(x_i(\theta_i)) + \theta_i = c$ and the number of motorists on road $i$ just depends on the toll $\theta_i$. As in Knight's case ($n = 1$), $x_i(\theta_i)$ is a strictly decreasing function of the one variable, $\theta_i$.

The toll setter $i$ ($i = 1, \ldots, n$) wants to maximize:

$$\pi_i(\theta_i) = \theta_i x_i(\theta_i) = x_i(\theta_i)(c - A_i(x_i(\theta_i)) \text{ with respect to } \theta_i. \tag{2}$$

The toll setter's maximand is $\pi_i(\theta_i) = R_i(x_i(\theta_i))$, which reaches a global maximum when the number of motorists on the congestible road is socially optimal because it is the planner's objective function from section 3.[10]

# 5 Assigning Property Rights Does Not Always Eliminate Inefficiency

The previous section extended Knight's claim that property rights can solve the market inefficiency to the case with $n$ congestible roads.

We demonstrated there that the maximand for the toll setters is equal to the social planner's minimand. In that case, revenue-maximizing tolls are equal to the "Pigouvian" taxes. Another useful way of understanding why toll-setting is efficient is to consider toll setters' optimization in this case. Suppose there are two toll roads. The toll setters maximize their revenue $\theta_i x_i$ for $i = 1, 2$ subject to the constraint that the total cost to the motorists inclusive of the toll equals the exogenous cost on the uncongestible road, $A_i(x_i) + \theta_i = c$. This constraint determines the rate at which motorists shift to and from a given congestible road for a given change in toll:

$$\begin{aligned} A_i'(x_i)dx_i + d\theta_i = dc &= 0 \\ \Rightarrow \frac{\partial x_i}{\partial \theta_i} &= \frac{-1}{A_i'(x_i)}. \end{aligned}$$

Substituting this into the first-order condition $\theta_i \frac{\partial x_i}{\partial \theta_i} + x_i = 0$ demonstrates the revenue-maximizing toll on each road equals the road-specific "Pigouvian" tax: $\theta_i^P = A_i'(x_i)x_i$ for $i = 1, 2$.

However, property rights cannot *always* solve this externality problem. Consider the case in which some motorists do not use the uncongestible road, because it either does not exist or is

[10] The details are the same as in the case of one congestible road and are explicitly given in Appendix A.

sufficiently slow. In this case, if a toll setter raises or lowers his toll, the number of motorists on the other congestible roads changes; no one switches to or from the uncongestible road: $dx_1 + dx_2 = 0$. Consider the first-order condition for the toll setter $\theta_i \frac{\partial x_i}{\partial \theta_i} + x_i = 0$ in conjunction with the condition derived by totally differentiating the equilibrium constraint from the second stage:

$$\begin{aligned} A_i'(x_i)dx_i + d\theta_i &= A_j'(x_j)dx_j \\ \Rightarrow \frac{\partial x_i}{\partial \theta_i} &= \frac{-1}{A_i'(x_i) + A_j'(x_j)}. \end{aligned} \tag{3}$$

The effect of changing the toll on the number of drivers differs from the earlier case because now it affects the level of drivers on the other congestible road. In this circumstance, the revenue maximizing toll is given by $\theta_i^* = x_i(A_1'(x_1) + A_2'(x_2))$. This pair of private tolls no longer leads to the efficient allocation of drivers unless $x_1^s A_2'(x_2^s) = x_2^s A_1'(x_1^s)$, which occurs when the roads are the same. The role of heterogeneity is discussed further in section 8. Therefore there is something important in having an uncongestible road that allows property rights to solve the market inefficiency.

Knight's elaboration of Pigou's example is frequently taken as evidence that the assignment of property rights insures efficiency. But Knight (1924, p. 591) refers to "competitive conditions" and even Demsetz (2010, p. 4) refers to "a competitive setting." Both seem to acknowledge that it is not the mere assignment of property rights that insures efficiency; indeed, both authors implicitly recognize that efficiency is unlikely to result from use of the price system unless competitive conditions prevail. But it is unclear what would constitute competitive conditions in Demsetz's case where both roads are to some extent congestible.

Suppose there is no uncongestible road but instead there are $M$ times as many motorists and $M$ times as many congestible roads of each type. We know from the current section that if $M = 1$, toll setters competing for the business of the motorists will not in equilibrium achieve an efficient allocation of motorists. In the Appendix B, we consider the case where $M$ grows without bound. We show that in this case the incentive for toll setters to deviate unilaterally from an efficiency-inducing toll profile disappears. This provides one concrete definition of competitive conditions. Intuitively, for any $M$, the efficient number of motorists on any road of a given type does not change. Moreover, for any profile of tolls, the number of motorists choosing a roadway of a given type does not change. What changes is the rate at which a toll setter on a road of a given type anticipates that he will lose customers if he marginally increases his toll. For small $M$, this rate of loss is dampened since every competing road is rendered less attractive as the motorists fleeing the increased toll add to its congestion. But as $M$ grows, the fleeing motorists locate on so many

different roads that the additional congestion imposed on any one of them becomes negligible. In the limit, therefore, it is *as if* the alternative to any individual toll setter's road is a completely uncongestible alternative.

A thick market provides the necessary "competitive conditions" (to use Demsetz's phrase) for property rights to eliminate the inefficiency because the effect of a change in toll in one road has increasingly less effect on the number of drivers on other roads until in the limit $dx_j/d\theta_i \to 0$. In this case the second term that produces the wedge in equation (3) goes to zero and efficiency is restored.

# 6 An Analysis of Duopoly Toll-Setting as the Cost of the Uncongestible Road Is Varied Exogenously

## 6.1 Overview

The analysis in section 4 implicitly assumed that the exogenous value of $c$ was so low that the motorists utilized all the uncongestible road. It concluded that private toll-setting leads to an efficient allocation. Define $\underline{c}$ as the highest cost for which the equilibrium has these characteristics. The analysis in section 5 implicitly assumed an exogenous value of $c$ so high that no motorists used the uncongestible road and the common cost of commuting on the two toll roads, including tolls, was strictly cheaper than the cost of using the uncongestible road. It concluded that, contrary to Knight's conclusion, private toll-setting by duopolists is inefficient. Define $\bar{c}$ ($> \underline{c}$) as the lowest cost for which the equilibrium has these characteristics.

In this section, we analyze the subgame perfect equilibrium of the duopoly toll-setting game more systematically, characterizing in the process the equilibria that arise in the overlooked intermediate interval where $c \in (\underline{c}, \bar{c})$. For $c$ in this intermediate interval, no motorist uses the uncongestible road but potential competition from that road leads to toll-setting equilibria where the common cost on the two roads equals the exogenous cost $c$. These equilibria need not be efficient, again contradicting Knight's conclusion; but potential competition from the uncongestible road may improve the social welfare that would be generated by the private duopoly toll-setting in the absence of an uncongestible road. We conclude the section by pointing out how provision of such potential competition may justify government intervention in markets with congestion.

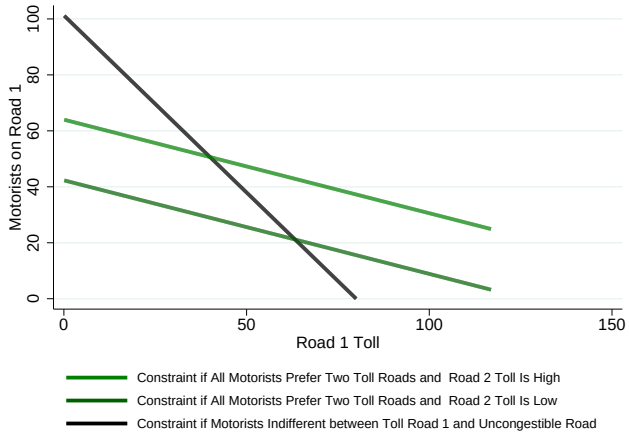
We begin by considering the best choice problem of each toll setter. We first derive the shape of the best replies to each other's toll and then describe how the Nash equilibria change as the exogenous cost $c$ is varied. To construct the best reply of toll setter 1 for any given $c$, we first determine, for every conjectured toll on road 2, how many motorists will use road 1 for each choice

of $\theta_1$, and hence the size of toll setter 1's revenue ($\theta_1 x_1$). Suppose, given the conjectured toll on road 2, that the toll on road 1 induced motorists to use all three roads. If toll setter 1 instead marginally reduced his toll, motorists would switch to his road until the cost of traveling on it returned to $c$. None of these additional motorists would come from toll road 2 since the conjectured toll on that road is, by assumption, unchanged and hence congestion on that road must also be unchanged in order for its overall cost to remain equal to $c$. Since $A_1(x_1) + \theta_1 = c$ throughout this range, $dx_1/d\theta_1 = -1/A_1'(x_1) < 0$.

Eventually, however, there will be no more motorists to attract from the uncongestible road. At that point, a further reduction in the toll on road 1 will attract motorists from road 2. On the interior of this region the two congestible roads have a common cost that is strictly smaller than the exogenous cost of commuting on the uncongestible road. Since $A_1(x_1) + \theta_1 = A_2(N - x_1) + \bar{\theta}_2 \leq c$ throughout this range, $dx_1/d\theta_1 = -1/[A_1'(x_1) + A_2'(N - x_1)] < 0$. So the budget constraint is kinked at the boundary between these two regions and has a strictly smaller magnitude to the left of the kink than to its right. Intuitively, a given reduction in $\theta_1$ will attract fewer motorists when they come from the other toll road rather than from the uncongestible road because the other toll road becomes more attractive as congestion on it declines.

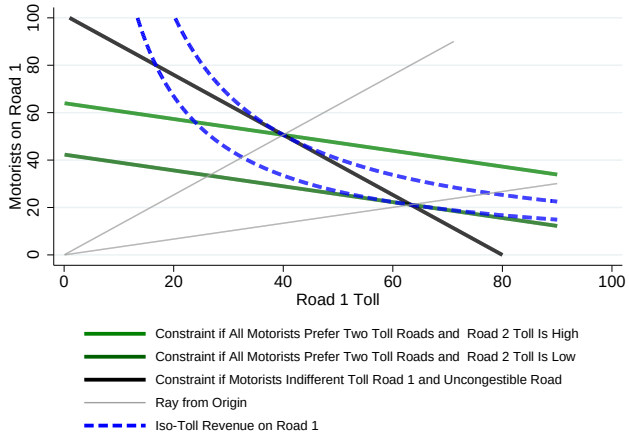
Assume $A_1(\cdot)$ is linear. Then, with $\theta_1$ on the horizontal axis and $x_1$ on the vertical axis, the budget constraint will be piecewise linear and will consist of a flatter downward-sloping segment when $\theta_1$ is sufficiently small and a steeper downward-sloping segment when $\theta_1$ is larger. If the toll conjectured for road 2 increases, the flatter portion of the budget constraint shifts out (still with slope $-1/[A_1' + A_2']$) since the same toll on road 1 must attract more motorists from road 2 because of its increased toll; the position of the steeper part of the budget constraint will not shift since, for any toll on road 1, the same level of congestion is required to raise the full cost on road 1 to $c$. Figure 6.1 displays in $(\theta_1, x_1)$ space the kinked budget constraint associated with a given toll on road 2 and shows how the flatter piece of it would shift up if the conjectured toll is higher.

Figure 6.1: Two Road and Uncongestible Road Constraints




Toll setter 1's isorevenue curves can be represented as a field of rectangular hyperbolas in the first quadrant satisfying the equation $\theta_1 x_1 = \text{constant}$. Since the slope of an indifference curve is $\text{MRS}(\theta_1, x_1) = -x_1/\theta_1$, any given ray from the origin cuts every indifference curve at the same slope. Two such rays are displayed in Figure 6.2. Ray OA cuts every indifference curve with the slope of the flatter part of the budget constraint $(-1/[A_1' + A_2'])$. Ray OB cuts every indifference curve with the slope of the steeper part of the budget constraint $(-1/A_1')$. Given the shape of the indifference curves and the budget constraint, the best reply for each conjectured toll on road 2 is unique.

Figure 6.2: The Optimal Choices for Toll Setter 1



We can now examine the *locus* of best choices $(\theta_1, x_1)$ as the toll conjectured on road 2 increases. If $\theta_2$ is very low, ray OA intersects the budget constraint on its flatter portion. This intersection point must be optimal since the indifference curve through it will be tangent to the flatter portion of the constraint. As the conjectured toll 2 increases, the flatter portion of the budget constraint intersects ray OA further from the origin. Hence, it is optimal for toll setter 1 to increase his toll.

Eventually the flatter portion of the budget constraint shifts out so far that it intersects ray OA on its steeper portion. The intersection of ray OA and the constraint is then no longer optimal since a northwest movement along the steeper part of the constraint is strictly preferred to that intersection point. The optimum in this case occurs at the kink point on the budget constraint. As the flatter portion shifts out, the kink point moves up the unchanging steeper segment of the budget constraints to the northwest. Hence, the optimal toll on road 1 *decreases* in response to the conjectured increase in $\theta_2$. Eventually, the kink point reaches B. If the road 2 toll is conjectured to be even higher, ray OB continues to cut the budget constraint at point B since the position of the steeper portion of the constraint does not change. In that case, the indifference curve through point B remains tangent to the constraint at B and point B remains optimal.

It is straightforward to deduce from this analysis the best reply of toll setter 1 to any conjectured $\theta_2$. As the conjectured toll on road 2 increases, the best toll on road 1 increases monotonically (strategic complements) until it reaches the horizontal component of point A. As the conjectured toll on road 2 increases further, the best toll on road 1 decreases monotonically (strategic substitutes) until the horizontal component of point B is reached. Any further marginal increases in the conjectured toll on road 2 do not affect the best toll on road 1 (strategic independence). If $A_i(x_i)$ for $i = 1, 2$ is linear, the best replies are piecewise linear. The slope on this first segment is $d\theta_2/d\theta_1 = 2$. Hence, the slope on the strategic-complements segment is constant and independent of $c$.[11] Points on the strategic-substitutes segment of toll setter 1's best reply must satisfy the following equations:

$$x_1 + x_2 = N \tag{4}$$

$$A_1(x_1) + \theta_1 = c \tag{5}$$

$$A_2(x_2) + \theta_2 = c. \tag{6}$$

If the congestion functions are linear, equations (5) and (6) will be linear and the system can be solved, for any given $c$, to determine the optimal $\theta_1$ for any given $\theta_2$. Differentiating, we conclude

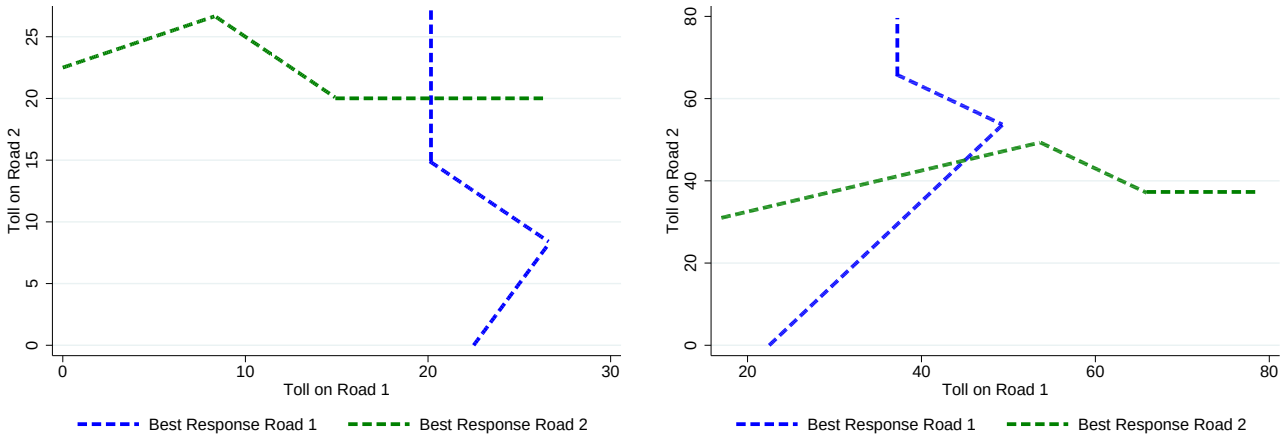
[11] The flatter piece of the budget constraint satisfies $A_1(x_1) + \theta_1 = A_2(N - x_1) + \theta_2$. The tangency condition is $1/[A_1'(x_1) + A_2'(x_2)] = x_1/\theta_1$. Assuming the two congestion functions are linear, we can eliminate $x_1$ using the tangency condition and can differentiate to obtain $d\theta_2/d\theta_1 = 2$; the slope of the other best reply in the strategic-complements region is $d\theta_2/d\theta_1 = 1/2$.

that on this segment $d\theta_2/d\theta_1 = -A_1'/A_2'$. Hence, the slope on this segment is also constant and is independent of $c$.[12] In the strategic-independence region, the best reply of toll setter 1 in $\theta_1 - \theta_2$ space is a vertical line.

The best reply of toll setter 2 to any conjectured $\theta_1$ can be deduced in the same manner. It will be piecewise linear, with the piece closest to the origin reflecting strategic complements, the next piece reflecting strategic substitutes, and the piece farthest from the origin reflecting strategic independence. Every point on the strategic-substitutes segment of toll setter 2's best reply must satisfy exactly the *same* three equations (4)-(6). Hence, both the strategic-substitutes segment for toll setter 1 and the strategic-substitutes segment for toll setter 2 must lie on the same line in $(\theta_1, \theta_2)$ space. But the two segments may be disjoint or may only partially overlap. If $A_i(x_i)$ for $i = 1, 2$ is linear, toll setter 2's best reply has a slope of $1/2$ in its strategic complements region, $A_1'/A_2'$ in its strategic substitutes region, and a slope of zero in its strategic independence region.

Figure 6.3 depicts the two best-reply functions. The two functions are drawn for some $c < \underline{c}$ and therefore intersect where each function reflects strategic independence. This is the region discussed by Knight and Pigou and in section 4. At the opposite extreme ($c > \bar{c}$) each function would intersect in its upward-sloping region. In this region, private sector toll-setting is inefficient as shown in section 5.

Figure 6.3: Best Replies When $c$ Is Low and When $c$ Is High




In both regions, the Nash equilibrium is unique. If $c < \underline{c}$, toll setter 1's best reply continues up vertically and toll setter 2's best reply continues to the right horizontally so they cannot cross a second time with one or both tolls larger. Moreover, on the interior of this region the two

[12]The characterization of the equilibria in the linear example is done in Appendix C.

strategic-substitutes segments must be disjoint. If $c > \bar{c}$, the best replies intersect in their respective strategic-complements segments. Their respective strategic-substitutes regions are again disjoint and so there can be no other Nash equilibrium.

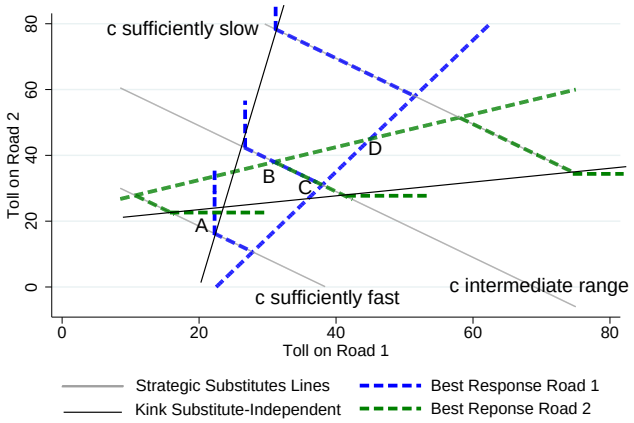
If $c \in (\underline{c}, \bar{c})$, the two downward-sloping pieces *partially* overlap. Hence, in this range, a given $c$ can generate a *continuum* of equilibria. Consider two equilibria associated with the same $c$. These can be visualized as lying on line segment AB in Figure 6.2. The equilibrium with the smaller toll on road 1 will have more motorists on road 1, a larger toll on road 2, and fewer motorists on road 2. In each equilibrium, the total cost on each road (inclusive of the tolls) will be $c$. Since the planner's optimization problem results in a unique number of motorists on each road, all (or all but one) of the Nash equilibria in the strategic-substitutes region are inefficient. Hence, Knight's belief that self-interested toll-setting by the private market must insure efficiency is mistaken. Thus, there remains an efficiency rationale for government intervention.

It is instructive to understand how the equilibrium changes as $c$ varies. If $c$ is lower, the steeper piece of the budget constraint will be closer to the origin but will still have the same slope. Consequently, this shifted-in steeper portion of the constraint, depicted in Figure 6.2, will intersect ray OB at a smaller $\theta_1$. Moreover, the smallest $\theta_2$ that generates an optimum where ray OB intersects the shifted-in budget constraint is also smaller. Finally, the shifted-in portion will intersect ray OA closer to the origin.

Hence, the new strategic-complements portion at first coincides with the previous strategic-complements portion but ends at a point closer to the origin. The strategic-substitutes segment is therefore closer to the origin and has the same slope as before; but the kink where the reaction function becomes vertical occurs at a smaller $\theta_1$ and a smaller $\theta_2$.

As for the best reply of toll setter 2, as $c$ is reduced, the new strategic-complements portion also coincides initially with the previous strategic-complements portion but ends closer to the origin. The strategic-substitutes segment has the same slope as before and is again a segment of the same line from which toll setter 1's strategic substitutes piece is extracted. The kink where the reaction function of toll setter 2 becomes horizontal occurs at a smaller $\theta_2$ and a smaller $\theta_1$. Figure 6.4 shows how the equilibrium tolls change as $c$ varies.

Figure 6.4: Best Replies When $c$ Varies from High to Low




Point A in Figure 6.4 depicts the equilibrium when the uncongestible road is fast, inducing an equilibrium in the strategic independence region as in section 4. The line segment BC depicts the continuum of equilibria when the uncongestible road's speed is in the intermediate region $c \in [\underline{c}, \bar{c}]$. Point D in Figure 6.4 depicts the strategic-complements region equilibrium from section 5 that occurs when the uncongestible road is sufficiently slow.

The next section discusses the welfare implications using two graphs. Figure 6.5 translates the equilibria found in Figure 6.4 into (total commute time, uncongestible road commute time) space. Figure 6.6 is a zoomed-in figure of Figure 6.4 focusing in on the intermediate range. Specifically, it illustrates how changes in line segment BC in Figure 6.4, due to changes in the commute time on the uncongestible road c, translate into different total commute times that may or may not be efficient.

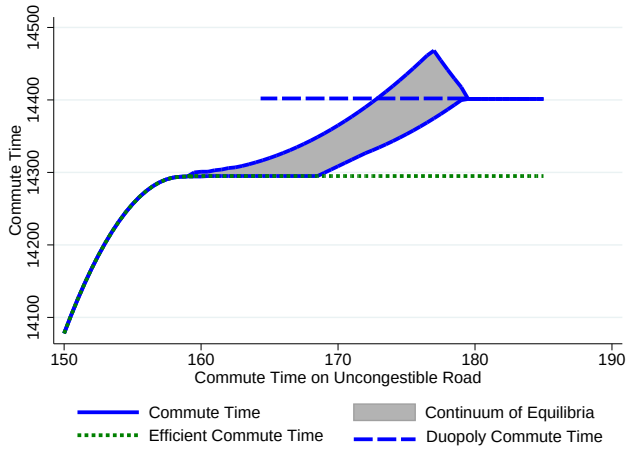
## 6.2 Welfare Implications: Pigouvian Taxes versus Private Toll-Setting

Figure 6.5 depicts the aggregate commute time (total man-hours lost commuting) in the efficient solution and the toll-setting equilibrium as the exogenous speed on the uncongestible road is changed. The aggregate commute time in the planning solution is strictly increasing in $c$ until $c = \underline{c}$ and is constant thereafter since the planner finds the uncongestible road too slow to utilize. Aggregate commute time in the toll-setting equilibrium coincides with that in the planning solution for $c \leq \underline{c}$. At the other extreme ($c > \bar{c}$), aggregate commute time is strictly higher than in the planning solution.

In the intermediate region, there is a continuum of equilibria. For some $c$, there exists a pair of

tolls that forms a Nash equilibrium and that achieves the aggregate commute time of the planning solution. But there are also equilibria with aggregate commute times even longer than in the solution with no uncongestible road.

Figure 6.5: Commute Time as Speed on Uncongestible Road Varies




If $c$ is in the intermediate third interval, some novel benefits of the uncongestible road appear. In this case, all of the motorists utilize the congestible roads, but nonetheless the total cost on the roads is equal to that on the uncongestible road. Therefore, there is a kink in the condition toll setters face. If they raise their toll, they push drivers onto the uncongestible road; in contrast, if they lower their toll, they pull drivers off of the other congestible roads.

Lower Toll:

$$\theta_i \frac{\partial x_i}{\partial \theta_i} + x_i > 0$$

$$A_i'(x_i)dx_i + d\theta_i = -dA_j'(x_j)dx_i \qquad \frac{\partial x_i}{\partial \theta_i} = \frac{-1}{(A_i'(x_i) + A_j'(x_j))} \tag{7}$$

Raise Toll:

$$\theta_i \frac{\partial x_i}{\partial \theta_i} + x_i < 0$$

$$A_i'(x_i)dx_i + d\theta_i = 0 \qquad \frac{\partial x_i}{\partial \theta_i} = \frac{-1}{A_i'(x_i)} \tag{8}$$

Nash Equilibrium Tolls:

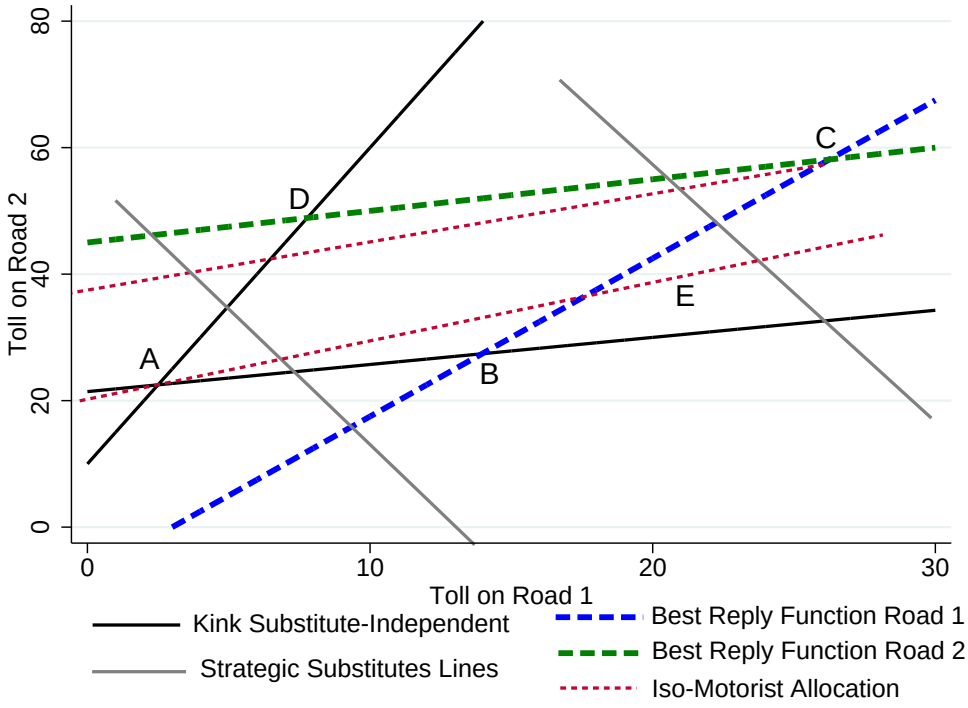
$$\theta_i \in \left[A_i'(x_i)x_i, (A_i'(x_i) + A_j'(x_j))x_i\right] \tag{9}$$

These conditions produce a set of toll-setting equilibria that can be supported when $c \in (\underline{c}, \bar{c})$. It can easily be shown that (i) for any $c \in (\underline{c}, \bar{c})$ there always exists at least one equilibrium; (ii) in any such equilibria, the uncongestible road is always empty; (iii) in any such equilibria, the total cost to motorists equals the time cost on the uncongestible road; and (iv) for any given $c \in (\underline{c}, \bar{c})$ there is no equilibrium such that the uncongestible road is ignored by the toll setters.

The total commute time in this scenario depends on the specific equilibrium. Figure 6.6 plots the continuum of equilibria in (road 1 toll, road 2 toll) space, zoomed in on a part of Figure 6.4. The continuum of equilibria is contained in the area ABCD, which is bounded on two sides by the strategic complement segment of each best reply and on the other two sides by the line that defines the kink between the strategic substitutes and strategic independence segment of each best reply. Within area ABCD, toll pairs lying on a line of slope 1 induce subgame-perfect equilibria with an unchanging number of motorists on each road. Two such lines are depicted in the diagram. As one moves from point A to point E, for example, each toll setter increases his toll by the same amount and therefore motorists have no incentive to switch to the other road. Hence, aggregate commute time remains minimized. However, toll pairs on this line closer to point E induce equlibria where the payoffs to motorists are reduced by exactly as much as payoffs to toll setters are increased. We have also drawn a line parallel to AE through point C, the situation with no uncongestible road. The allocation of motorists is unchanging in all the equilbria induced by toll pairs lying on this line since it slopes up at 45°. Hence they all have the same aggregate commute time as in the equilibrium where no uncongestible road is available. Now consider any point lying strictly north of a point on this line but still within area ABCD. Since toll 2 is higher and and toll 1 is unchanged, this equilibrium will have more motorists on road 1 and fewer on road 2 than at point C. In any such equilibrium the aggregate commute time is even longer than if there were no uncongestible road. This explains why some of the shaded region in Figure 6.5 lies above the higher of the two horizontal lines.[13]

[13] In the case where both congestible roads are symmetric points A and C lie on the 45° line from the origin. In this case, the efficient allocation is feasible for any $c \in [\underline{c}, \bar{c}]$.

Figure 6.6: Continuum of Equilibria with Duopoly and Efficient Allocations




To summarize, Knight's contention that government intervention is unnecessary in congestion problems because private toll setters can be relied upon to set access fees on the various contestable resources at the same levels as the government would is mistaken. As Figure 6.5 makes clear, his novel contention is correct for $c \leq \underline{c}$ and is surprising given the first welfare theorem. But it is not generally true. In particular, it is false if $c > \underline{c}$.

When "competitive conditions" exist assigning property rights *does* restore efficiency. As we have seen, in this context "competitive conditions" mean either that there exists a sufficiently attractive uncongestible option or that there exist enough competing toll setters (formalized as a replica economy). However, there are many real-world examples of congestion where such "competitive conditions" do not exist. In these cases, assigning property rights alone will, contrary to Knight's contention, fail to restore efficiency and there is a role for the government. As Pigou originally asserted, the government could restore efficiency by imposing road-specific taxes. Our analysis suggests, however, a new type of government intervention: the government can provide an uncongestible option.

# 7 Lessons for Government Intervention

Suppose there were several congestible routes connecting Washington, D.C., and New York. If competing private toll setters required payment for access to each route, competition between the toll setters would not minimize aggregate time wasted commuting. Suppose, however, that the government provided a slow but uncongestible alternative—a train. A train is uncongestible since it can be any length. If the time between one train and the next was sufficiently short, the average commute time on the uncongestible road would be fast enough that the toll setters would set their tolls at the Pigouvian levels. If trains were scheduled less frequently, however, fewer commuters would use them. Suppose the frequency was diminished just to the point where no commuter used the train: $c = \underline{c}$.

In this case, the toll-setting equilibrium is still efficient. What stops either toll setter from unilaterally raising his toll is his recognition that he would lose too many motorists to the uncongestible road for this to be profitable. Suppose now that the time between trains became longer so that on average motorists would lose $\Delta$ more in wages by taking a train. It is our hypothesis that each toll setter would raise his toll by $\Delta$, reasoning that he could safely do so without losing any motorists to the uncongestible alternative provided the other toll setter reacted in the same way. That is, the cost of using the uncongestible road serves as a focal point for each toll setter. If this hypothesis is correct, then as we raise $c$ from $\underline{c}$ toward $\bar{c}$, the toll setters will respond by raising their tolls in a way that maintains aggregate commute time at the efficient level. The government would incur fewer costs since it would be running fewer trains and the motorists would pay more in tolls to get to work at the same time as before.[14] Toll setters would enjoy higher revenues. The potential competition from the government train would maintain efficiency. Despite this, no commuter would ride the train.

This analysis demonstrates an attractive alternative to Pigouvian taxes, in which the government provides an uncongestible outside option to discipline the market. With perfect information the Pigouvian tax always achieves efficiency. However, there is disagreement about the practicality of Pigouvian interventions: "Virtually every author points out that we do not know how to calculate the ideal Pigouvian tax or subsidy levels in practice, but because the point is rather obvious rarely is much made of it" (Baumol and Oates (1971)). Providing an uncongestible outside option

[14] If the two roads are equally congestible, even a cost increase of $\bar{c} - \underline{c}$ can induce an efficient subgame perfect equilibrium with each toll setter raising his toll by $\Delta = \bar{c} - \underline{c}$. If the roads are asymmetric, however, there will exist a maximum $c^* < \bar{c}$ beyond which the equilibrium is necessarily inefficient. If the two toll setters raised their tolls by $\bar{c} - c^*$, at least one of the toll setters would have an incentive to steal his rival's customers by cutting his toll. For example, point E on line AE in Figure 6.6 is not an equilibrium because toll setter 1 has an incentive to lower his toll since his best reply passes to the left of point E. By doing so, he expects that the number of motorists he can attract will rise by a larger percentage than the percentage reduction in his toll.

requires far less information and can produce the efficient allocation even when the outside option is unused. Further, there are circumstances in which the government providing an outside option may be the preferable option, such as, providing a low-cost health care alternative, providing a national lab to employee engineers, or providing alternative work to fishermen.

# 8 Comparative Statics

This section discusses the impact of (1) heterogeneity in congestion across roads, (2) the number of motorists, and (3) the number of roads on the ability of the government to use the proposed intervention of providing an uncongestible alternative. The discussion on heterogeneity demonstrates under what circumstances government intervention is most useful—markets characterized by similar or dissimilar competitors. The analysis discussing the impact of changes in the number of motorists is important in both the short and long run as the number of motorists is likely to differ within a given day, during rush hour, and across years. The analysis discussing the number of roads is important in thinking about the effectiveness of the government's uncongestible option in different industries and dynamic industries where the number of firms is likely to change.

## 8.1 Impact of Heterogeneity across Roads

The efficient toll depends on the congestion on the road and therefore may differ as the congestion differs. As the congestion on a road increases, $x_i A_i'(x_i)$, the efficient toll increases causing the efficient allocation of motorists to have fewer motorists on road $i$ and weakly more motorists on road $j$, because the allocation of motorists depends on the tolls. The allocation of motorists is not uniquely determined by a pair of tolls, in the strategic-complements case, but by the difference in tolls. Therefore, to consider the effect of heterogeneity on the allocation of motorists, we consider the difference in equilibrium tolls in the strategic-complements case without an uncongestible road.

$$\begin{aligned} \theta_1^* &= x_1 A_1'(x_1) + x_1 A_2'(x_2) \\ \theta_2^* &= x_2 A_2'(x_2) + x_2 A_1'(x_1) \\ \theta_1^* - \theta_2^* &= \theta_1^P - \theta_2^P + \underbrace{x_1 A_2'(x_2) - x_2 A_1'(x_1)}_{\text{Wedge}} \end{aligned} \tag{10}$$

When the difference in tolls is equal to the difference in "Pigouvian" tolls the motorist allocation is efficient, even if the tolls do not equal the "Pigouvian" tolls. Equation (10) demonstrates that the equilibrium difference in tolls, in the strategic-complements case without an uncongestible road, is equal to the difference in "Pigouvian" tolls plus a wedge term. Consider the case where both

congestible roads are identical. In this case the wedge is zero and the allocation of motorists is efficient. This occurs because even though the equilibrium tolls are higher than the "Pigouvian" tolls the increase in toll is the same for both roads.

As the roads become more heterogeneous the wedge term increases in magnitude causing the allocation of motorists to become farther from the efficient allocation. This suggests the benefits from government intervention are highest in cases where congestion is heterogeneous. This analysis is consistent with the literature on dynamic city growth which demonstrates heterogeneity across cities has important implications for the effects of tax and zoning laws on the allocation of population across cities (Seegert (2011)).

### 8.2 Impact on Changing the Number of Motorists: Rush Hour

Consider the impact on the equilibria found in section 6 as the number of motorists increases but everything else is held constant—importantly the number of roads and the speed of the uncongestible road $c$. Equilibria exist in three regions defined by the threshold levels $\underline{c}$ and $\bar{c}$, strategic independence ($c < \underline{c}$), substitutes($c \in [\underline{c}, \bar{c}]$), and complements ($c > \bar{c}$). Increasing the number of motorists changes these threshold levels and therefore the set of equilibria. The threshold levels are defined as the level of $c$ equal to the total cost to motorists on a congestible road where the equilibrium toll is set independently or strategically where tolls are strategic complements.

$$\begin{aligned} \underline{c} &= A_i(x_i) + \theta_i \quad \text{when } \theta_i = A_i'(x_i)x_i \\ \bar{c} &= A_i(x_i) + \theta_i \quad \text{when } \theta_i = (A_i'(x_i) + A_j'(x_j))x_i \end{aligned} \tag{11}$$

The effect of increasing the number of motorists on these threshold levels is determined by totally differentiating the conditions and rearranging.

$$\frac{d\underline{c}}{dN} = (2A_i'(x_i) + A_i''(x_i)x_i)\frac{\partial x_i}{\partial N} > 0$$

$$\frac{d\bar{c}}{dN} = (2A_i'(x_i) + A_j'(x_j) + A_i''(x_i)x_i)\frac{\partial x_i}{\partial N} + A_j''(x_j)x_i\frac{\partial x_j}{\partial N} > 0$$

The threshold levels both increase as the number of motorists increases. This implies that for a given speed of the uncongestible road as the number of motorists increases the equilibria shift

from strategic complements to strategic substitutes to strategic independence. If the increase in motorists is enough to shift the set of equilibria between these regions then it is possible for a government intervention that did not discipline the market at all with fewer motorists to discipline the market to the efficient allocation with more motorists. Even if the change is not large enough to discipline the market to the efficient allocation it may still increase efficiency.

If the change in motorists is not large enough to change the set of equilibria then efficiency may increase in the intermediate region if the uncongestible road disciplines the markets to a different set of equilibria that may be more efficient. The tolls will remain efficient in the strategic-independence region and remain inefficient in the strategic-complements regions. Within the context of rush hour, where there is a spike in the number of motorists on the roads, this analysis suggests the government intervention may become more effective, and will not become less effective, at disciplining the market during rush hour especially if the change is large.

### 8.3 Impact on Changing the Number of Roads

Now consider the impact of the equilibria when the number of roads increases by replicating the system of roads, but not the number of motorists, $M$ times. The effect of replicating the number of roads on the threshold levels is given by totally differentiating the constraints in equation (11).

$$\frac{d\underline{c}}{dM} = (2A_i'(x_i) + A_i''(x_i)x_i)\frac{\partial x_i}{\partial M} < 0$$

$$\frac{d\bar{c}}{dM} = (2A_i'(x_i) + A_j'(x_j) + A_i''(x_i)x_i)\frac{\partial x_i}{\partial M} + A_j''(x_j)x_i\frac{\partial x_j}{\partial M} < 0$$

The threshold levels decrease as the number of roads increases. Intuitively, as the number of roads increases the time cost of commuting on any given road decreases as motorists are spread across more roads. Therefore, the threshold time cost for some motorists to be on the uncongestible road, $\underline{c}$, must be lower than before. As the number of roads increases, and the speed on the uncongestible road stays the same, eventually the uncongestible road that was being used, and disciplining the market, no longer is fast enough to be used in equilibrium. As the number of roads increases, the uncongestible road becomes less effective at disciplining the market.

As the number of roads increases, the government intervention becomes less effective but the benefit from disciplining the market also decreases. As demonstrated previously as the number of replications increases to infinity the inefficiency caused by the market decreases to zero. However, if we consider the case in which the number of competitors decreases, for instance if one road

is unused because of construction, then the government intervention becomes more effective. If market structure changes to consolidate the number of competitors in the market the government intervention automatically disciplines the market more, counteracting some of the possible negative impact of market power.

While these seem like two separate concerns it turns out that both can be summarized by the level of congestion in the market. Congestion in the market can be thought of as the level of negative externality motorists impose on other motorists, $x_i A_i'(x_i)$. As the number of motorists increases, either during rush hour or across many years, the congestion in the market increases. Conversely, as the number of competing roads increases, holding the number of motorists fixed, the congestion decreases.

These comparative statistics reinforce the usefulness of this new government intervention. First, demonstrated through the example of an increase in motorists, we see providing an uncongestible alternative disciplines the market more when the number of motorists increases and congestion is high. Second, as industries evolve, changing the number of competitors, roads in this example, we see the government intervention disciplines the market most when the level of competition is low, demonstrated through the example of an increase in roads. In both scenarios the government intervention's effectiveness at disciplining the market increases precisely when doing so is most beneficial—when congestion is high and competition is low. Third, the ability of this government intervention to adapt to these changing scenarios is automatic, requiring no changes in law or additional information on how congestion has changed.

# 9 Conclusion

Since congestion problems abound, our analysis applies to allocation problems unrelated to highway commuting. To remind readers of this, we close with an application that is seemingly very different. Suppose $N$ medical researchers may be assigned either to the National Institute of Health (NIH) or to independent research projects, each of which seeks a cure for one of $n < N$ unrelated diseases. Contests to stimulate solutions to such problems are becoming ubiquitous. Assume researchers can enter at most one contest and each researcher is equally likely to find a cure for a particular disease (although some diseases are more intractable than others). Since a researcher's chance of winning a prize in contest $i$ is strictly decreasing in the number of other researchers with whom he is competing, this is a congestion problem. Consider how a social planner would allocate researchers across these $n + 1$ activities to maximize expected social surplus. The planner could achieve this allocation by setting $n$ project-specific "Pigouvian" prizes (the analog of $n$ highway-specific Pigouvian tolls) to induce researchers (the analog of the motorists) to allocate themselves

efficiently across the $n$ contests (the analog of the congestible roads) and NIH (the analog of the uncongestible road). Knight's insight suggests that it is unnecessary for the government to offer Pigouvian prizes since independent prize setters would always find it in their self-interest to offer them. As in the case of highway tolls, however, this conclusion is incorrect. Prize-setting will not in general result in an efficient allocation of researchers across contests. And when the allocation is inefficient, there is again a role for the government to intervene.[15]

[15]We elaborate on this application in Appendix D.

## References


**Anderson, Michael L**, "Subways, strikes, and slowdowns: The impacts of public transit on traffic congestion," Technical Report, National Bureau of Economic Research 2013.

**Arnott, Richard, Andre De Palma, and Robin Lindsey**, "Schedule delay and departure time decisions with heterogeneous commuters," *Transportation Research Record*, 1988, *1197*, 56–67.

**___, André De Palma, and Robin Lindsey**, "Economics of a bottleneck," *Journal of urban Economics*, 1990, *27* (1), 111–130.

**___, Andre De Palma, and Robin Lindsey**, "A temporal and spatial equilibrium analysis of commuter parking," *Journal of public economics*, 1991, *45* (3), 301–335.

**___, Andre de Palma, and Robin Lindsey**, "Route choice with heterogeneous drivers and group-specific congestion costs," *Regional Science and Urban Economics*, 1992, *22* (1), 71–102.

**___, Andre De Palma, and Robin Lindsey**, "A structural model of peak-period congestion: A traffic bottleneck with elastic demand," *The American Economic Review*, 1993, pp. 161–179.

**___, André de Palma, and Robin Lindsey**, "The welfare effects of congestion tolls with heterogeneous commuters," *Journal of Transport Economics and Policy*, 1994, pp. 139–161.

**___, Andre de Palma, and Robin Lindsey**, "Recent Developments in the Bottleneck Model," *Roach Pricing, Traffic Congestion and the Environment: Issues of Efficiency and Social Feasibility*, 1998.

**Berglas, Eitan**, "On the theory of clubs," *The American Economic Review*, 1976, pp. 116–121.

___ , "The market provision of club goods once again," *Journal of Public Economics*, 1981, *15* (3), 389–393.

___ **and David Pines**, "Clubs, local public goods and transportation models: A synthesis," *Journal of Public Economics*, 1981, *15* (2), 141–162.

**___, Elhanan Helpman, and David Pines**, "The economic theory of clubs: Some clarifications," *Economics Letters*, 1982, *10* (3), 343–348.

**Block, Walter**, "Free market transportation: Denationalizing the roads," *Journal of libertarian studies*, 1979, *3* (2), 209–238.

___ , "Congestion and road pricing," *Journal of Libertarian Studies*, 1980, *4* (3), 299–330.

___ , "Road socialism," *International Journal of Value-Based Management*, 1996, *9* (2), 195–207.

**Boadway, Robin**, "A note on the market provision of club goods," *Journal of Public Economics*, 1980, *13* (1), 131–137.

**Brennan, T., M. Macauley, and K. Whitefoot**, "Prizes, Patents, and Technology Procurement: A Proposed Analytical Framework," *Resources for the Future DP 11-21*, 2011.

**Buchanan, James M**, "Private ownership and common usage: The road case re-examined," *Southern Economic Journal*, 1956, pp. 305–316.

___ , "An economic theory of clubs," *Economica*, 1965, pp. 1–14.

**Button, Kenneth J**, "How should the revenues from congestion pricing be spent?," *Street Smart: Competition, Entrepreneurship, and the Future of Roads*, 2006, pp. 225–242.

**Cadin, Michelle S and Walter Block**, "Privatize Public Highways," *Freeman-New Yrok-Foundation for Economic Education*, 1997, *47*, 96–97.

**Cheung, Steven N. S.**, "The Fable of the Bees: An Economic Investigation," *Journal of Law and Economics.*, 1973, *16*, 11.

**Coase, Ronald Harry**, "Problem of social cost, the," *Jl & econ.*, 1960, *3*, 1.

**Crocker, Thomas D**, "The structuring of atmospheric pollution control systems," *The economics of air pollution*, 1966, pp. 61–86.

**Dales, JH**, "Pollution, Property, And Prices," *Toronto: University of Toronto Press*, 1968.

**Demsetz, H.**, "Fallacies in the economic Doctrine of Externalities," *unpublished discussion paper*, 2010.

**Dupuit, J**, "De la Mesure de 1'Utilitie des Travaux publics'," in "Annales des Fonts et Chaussees," Vol. 8 1844.

**Edelson, Noel M**, "Congestion tolls under monopoly," *The American Economic Review*, 1971, pp. 873–882.

___ **and David K Hilderbrand**, "Congestion tolls for Poisson queuing processes," *Econometrica: Journal of the Econometric Society*, 1975, pp. 81–92.

**Edgeworth, Francis Ysidro**, *Mathematical psychics: An essay on the application of mathematics to the moral sciences*, C. Keagann Paul, 1881.

**Ellickson, Bryan, Birgit Grodal, Suzanne Scotchmer, and William R Zame**, "Clubs and the Market," *Econometrica*, 1999, *67* (5), 1185–1217.

**Fielding, Gordon J and Daniel B Klein**, "How to franchise highways," *Journal of Transport Economics and Policy*, 1993, pp. 113–130.

**Foldvary, Fred E**, "Streets as private-sector public goods," 2006.

**Henderson, J Vernon**, "Road congestion: a reconsideration of pricing theory," *Journal of Urban Economics*, 1974, *1* (3), 346–365.

**Hillman, Arye L and Peter L Swan**, "Participation rules for Pareto-optimal clubs," *Journal of Public Economics*, 1983, *20* (1), 55–76.

**Hughes, Jonathan E and Daniel Kaffine**, "When is Encouraging Consumption of Common Property Second Best? Sorting, Congestion and Entry in the Commons," Technical Report 2013.

**Klein, Daniel A and John D Majewski**, "Americas Toll Road Heritage: The Achievements of Private Initiative in the Nineteenth Century," 2006.

**Knight, F.H.**, "Some fallacies in the interpretation of social cost," *The Quarterly Journal of Economics*, 1924, *38* (4), 582–606.

**Kremer, M. and R. Glennerster**, *Strong medicine: Creating incentives for pharmaceutical research on neglected diseases*, Princeton University Press,, 2004.

**Lindsey, Robin**, "De Economists Reach a Conclusion?," *Econ Journal Watch*, 2006, *3* (2), 292–379.

**Malueg, David A**, "Welfare consequences of emission credit trading programs," *Journal of Environmental Economics and management*, 1990, *18* (1), 66–77.

**Mills, David E**, "Ownership arrangements and congestion-prone facilities," *The American Economic Review*, 1981, pp. 493–502.

**Parry, Ian WH and Kenneth A Small**, "Should urban transit subsidies be reduced?," *The American Economic Review*, 2009, *99* (3), 700–724.

**Pigou, A.C.**, *Welfare Economics*, London: Macmillan, 1920.

**Roth, Gabriel**, "Why involve the private sector in the provision of public roads," *Street Smart: Competition, Entrepreneurship and the Future of Roads*, 2006, pp. 3–21.

**Rothbard, Murray Newton**, *For a new liberty: The libertarian manifesto*, Ludwig von Mises Institute, 1978.

**Salant, SW.**, "Book Review: Strong Medicine: Creating Incentives for Pharmaceutical Research on Neglected Diseases," *Economic Development and Cultural Change*, 2007, *55* (4), 848.

**Salant, SW**, "Michael Kremer and Rachel Glennerster, Strong Medicine: Creating Incentives for Pharmaceutical Research on Neglected Diseases," *Economic Development and Cultural Change*, 2007, *55* (4), 848.

**Sandler, Todd and John T Tschirhart**, "The economic theory of clubs: An evaluative survey," *Journal of Economic Literature*, 1980, pp. 1481–1521.

**Scotchmer, Suzanne**, "Public goods and the invisible hand," *Modern Public Finance*, 1994, pp. 93–125.

___ **and Myrna Holtz Wooders**, "Competitive equilibrium and the core in club economies with anonymous crowding," *Journal of Public Economics*, 1987, *34* (2), 159–173.

**Seegert, N.**, "Barriers to Migration in a System of Cities," *unpublished discussion paper*, 2011.

**Small, Kenneth A and Jia Yan**, "The value of value pricing of roads: Second-best pricing and product differentiation," *Journal of Urban Economics*, 2001, *49* (2), 310–336.

**Stavins, Robert N**, "Transaction costs and tradeable permits," *Journal of environmental economics and management*, 1995, *29* (2), 133–148.

**Verhoef, Erik T, Peter Nijkamp, and Piet Rietveld**, "The social feasibility of road pricing: a case study for the Randstad area," *Journal of Transport Economics and Policy*, 1997, pp. 255–276.

**Vickrey, William S**, "Pricing in urban and suburban transport," *The American Economic Review*, 1963, pp. 452–465.

___ , "Congestion theory and transport investment," *The American Economic Review*, 1969, pp. 251–260.

**Walters, Alan A**, "The theory and measurement of private and social cost of highway congestion," *Econometrica: Journal of the Econometric Society*, 1961, pp. 676–699.

**Winston, Clifford and Jia Yan**, "Can privatization of US highways improve motorists' welfare?," *Journal of Public Economics*, 2011, *95* (7), 993–1005.

# Appendix A Global Maximum With "Sufficiently Fast" Uncongestible Road

The assumption that it is never advantageous for the toll setter to drive *every* motorist away from his road or attract every the motorist away from the uncongestible road holds for an uncongetible road that is "sufficiently fast." In section 6 the optimization is done in generality for all values of the uncongestible road. The maximum point found assuming a "sufficiently fast" uncongestible road is the global maximum. To show this consider the first order condition in the case where there are no motorists on the uncongestible road. In this case $\partial x_1/\partial\theta_1 = -1/(A_1'(x_1(\theta_1)) + A_2'(x_2(\theta_2))) > -1/A_1'(x_1(\theta_1))$. Therefore given the necessary condition for a maximum $x_1 + \theta_1\partial x_1/\partial\theta_1 = 0$ it is straightforward to show that a global deviation such that the uncongestible road is empty fails the necessary condition for a maximum, specifically $x_1^D + \theta_1^D\partial x_1/\partial\theta_1 > 0$.

$$
\begin{aligned}
x_1^D > x_1^s \quad & \text{Def. deviation} \\
x_1^s = \theta_1^s/(A_1'(x_1(\theta_1^s)) \quad & \text{FOC} \\
\theta_1^s/(A_1'(x_1(\theta_1^s)) > \theta_1^D/A_1'(x_1(\theta_1)) \quad & \text{know } \theta_1^D < \theta_1^s \\
\theta_1^D/A_1'(x_1^s(\theta_1)) > \theta_1^D/(A_1'(x_1^D(\theta_1)) + A_2'(x_2^D(\theta_2))) \quad & \text{know } A_1'(x_1(\theta_1)) + A_2'(x_2(\theta_2) > A_1'(x_1(\theta_1^s) \\
\Rightarrow \quad x_1^D > \theta_1^D/(A_1'(x_1^D(\theta_1)) + A_2'(x_2^D(\theta_2))) \quad & \Rightarrow x_1^D + \theta_1^D\partial x_1/\partial\theta_1 > 0
\end{aligned}
$$

# Appendix B Efficiency Restored in the Toll-Setting Game under "Competitive Conditions"

Suppose instead of $N$ motorists spread across $n$ motorways congestible to different extents there are $MN$ motorists spread across $n$ types of congestible roads, with $M$ identical roads of each type. We wish to show that as $M \to \infty$, the efficient allocation of motorists can be supported as a subgame-perfect equilibrium of the toll-setting game.[16]

The efficient solution in this case uniquely solves the following $n+1$ equations defining $\{x_i^s\}_{i=1}^n$ and $\lambda^s : A_i(x_i^s) + x_i^s A_i'(x_i^s) = \lambda^s,\ i = 1, \ldots, n$ and $\sum_{i=1}^n x_i^s = N$, where $\lambda^s$ denotes the value of the Lagrangean multiplier at the social optimum.

Define the Pigouvian tolls as $\theta_i^P = x_i^s A_i'(x_i^s)$ for $i = 1, \ldots, n$. Suppose at the first stage of the two-stage game every player but one type $i$ chooses the Pigouvian toll while this one remaining type $i$ player chooses some $\theta_i \in [\theta_i{}^{\min}, \theta_i{}^{\max}]$. Define these minimum and maximum tolls follows: $\theta_1^{\max} = \mu - A_1(0)$ and $\theta_1^{\min} = \mu - A_1(MN)$ where $\mu$ in the first condition is equal to $min\{A_j(0) + \theta_j^P\}$ and in the second equal to the common cost given by $A_j(x_j) + \theta_j^P$. Then in the Nash equilibrium of the second stage, the motorists will allocate themselves so the full cost (the toll plus the lost wages) is the same on every motorway. Denote this common cost as $\mu$. Then the following $n+2$

[16] Seegert (2011) demonstrates the efficiency of this system in the limiting case in the context of a system of cities.

equations define the $n+2$ variables $x_i^d, \{x_j\}_{j=1}^n,$ and $\mu$

$$\begin{aligned} A_i(x_i^d) + \theta_i &= \mu && \text{(B.1)} \\ A_j(x_j) + x_j^s A_j'(x_j^s) &= \mu \text{ for } j = 1, \dots n && \text{(B.2)} \\ x_i^d + (M-1)x_i + M\sum_{j \neq i} x_j &= N. && \text{(B.3)} \end{aligned}$$

Anticipating these second-stage responses and conjecturing that the other tolls setters will maintain their tolls at the Pigouvian level, the deviating type i player can set any toll $\theta_i$ in the closed interval. A marginal increase in his toll would then cause his profits to increase at the following rate:

$$\frac{\partial \pi_i}{\partial \theta_i} = \frac{\partial\big(\theta_i x_i^d(\theta_i)\big)}{\partial \theta_i} = x_i^d + \theta_i \frac{\partial x_i}{\partial \theta_i}. \tag{B.4}$$

Let $\theta_i = \epsilon_i + \theta_i^P = \epsilon_i + x_i^s A_i'(x_i^s)$. Differentiating (7) we obtain:

$$\frac{\partial x_i}{\partial \theta_i} = \left(\frac{\partial \mu}{\partial \theta_i} - 1\right) \frac{1}{A_i'(x_i^d)}. \tag{B.5}$$

Substituting for $\theta_i$ and recognizing that as $M \to \infty$, $\frac{\partial \mu}{\partial \theta_i} \to 0$, we conclude that

$$\frac{\partial \pi_i}{\partial \theta_i} = \frac{x_i^d A_i'(x_i^d) - x_i^s A_i'(x_i^s) - \epsilon_i}{A_i'(x_i^d)} \tag{B.6}$$

If $\epsilon = 0$, then the toll under consideration is the Pigouvian toll. By definition, $x_i^d = x_i^s$. So the expression in (12) equals zero. If $\epsilon > 0$, then the toll exceeds the Pigouvian level ($\theta_i > \theta_i^P$), less than the efficient number of motorists take the route ($x_i^d < x_i^s$), and the expression in (12) is strictly negative. Finally, if $\epsilon < 0$, then the toll is smaller than the Pigouvian level ($\theta_i < \theta_i^P$), more than the efficient number of motorists take the route ($x_i^d > x_i^s$), and the expression in (12) is strictly positive.

It follows that for any toll $\theta_i$ in the closed interval, the payoff of player $i$ is single-peaked at the Pigouvian toll. Moreover, any toll so high that no one uses the motorway or so low that raising it would not alter the number of motorists using his route is clearly suboptimal for toll setter $i$. But these arguments apply to *every* toll setter. Hence, in the limit no toll setter has a strict incentive to deviate unilaterally from the profile of Pigouvian tolls even when there is no uncongestible motorway.

## Appendix C Linear Example

Suppose the congestion on the roads are given by $A_i(x_i) = a_i x_i + b_i$. The best response functions are piecewise linear consisting of three parts; strategic complements, strategic substitutes, and strategic independence. Figure 6.2 demonstrates the optimal choices for toll setter 1 as the toll on the road 2 changes. When the toll on road 2 is low the optimal point is given by the tangency point between the iso-toll revenue curve and the constraint if all motorists prefer two toll roads. In this range the best reply function for toll 1 is a strategic complement to the toll on road 2.

When the toll on road 2 is an intermediate level, such that in Figure 6.2 the flatter constraint is between the two depicted, the optimal toll on road 1 is given by the kink point of the steeper and flatter constraints. In this range the best reply function for toll 1 is a strategic substitute with the toll on road 2. Finally, if the toll on road 2 is large, such that the flatter constraint is above the constraints drawn in Figure 6.2, then the optimal point is given by point B, the tangency between the iso-toll revenue on road 1 and the steeper constraint characterized by all motorists indifferent between toll road 1 and the uncongestible road. In this range the best reply function is strategically independent of the toll on road 2.

Deriving Strategic Complements: Low Values $\theta_2$.

$$\max \ \theta_1 x_1$$

$$\text{subject to } \ a_1 x_1 + b_1 + \theta_1 = a_2 x_2 + b_2 + \theta_2 \ \ N = x_1 + x_2$$

First-order condition

$$\frac{-2\theta_1}{a_1 + a_2} + \frac{a_2 N}{a_1 + a_2} + \frac{b_2 \theta_2 - b_1}{a_1 + a_2} = 0$$

$$\theta_1^{BR} = \frac{a_2 N + b_2 + \theta_2 - b_1}{2}$$

$$\theta_2 = 2\theta_1^{BR} + b_1 - a_2 N - b_2$$

Deriving Strategic Substitutes: Intermediate Values $\theta_2$.

The best reply function for intermediate values of $\theta_2$ is characterized by the point at which motorists are indifferent between being on toll road 1, toll road 2, and the uncongestible road but all motorists are on the two toll roads.

$$\begin{aligned} a_1 x_1 + b_1 + \theta_1 &= a_2 x_2 + b_2 + \theta_2 \\ a_1 x_1 + b_1 + \theta_1 &= c \\ N &= x_1 + x_2 \end{aligned}$$

Combining these three constraints produces:

$$\theta_2 c \frac{a_1 + a_2}{a_1} - b_1 \frac{a_2}{a_1} - \frac{a_2}{a_1}\theta_1 - b_2 - a_2 N.$$

In this range the best reply function for toll road 1 and toll road 2 are characterized by the same three conditions, thus lie on the same line. The best reply functions of toll road 1 and toll road 2 can overlap or be disjoint in this region as depicted in Figure 6.4.

Deriving Strategic Independence: High Values $\theta_2$.

The best reply function in this region is characterized by point $B$ in Figure 6.2. The best reply function can be derived from maximizing toll revenue given the constraint, $a_1 x_1 + b_1 + \theta_1 = c$. This leads to the best reply function,

$$\theta_1^{BR} = \frac{c - b_1}{2}.$$

Reaction Function Road 1
Strategic Substitutes: $\theta_2 = c\frac{a_1+a_2}{a_1} - b_1\frac{a_2}{a_1} - \frac{a_2}{a_1}\theta_1 - b_2 - a_2N$
Strategic Complements: $\theta_2 = 2\theta_1 + b_1 - a_2N - b_2$
Strategic Independence: $\theta_1 = \frac{c-b_1}{2}$

Reaction Function Road 2
Strategic Substitutes: $\theta_2 = c\frac{a_1+a_2}{a_1} - b_1\frac{a_2}{a_1} - \frac{a_2}{a_1}\theta_1 - b_2 - a_2N$
Strategic Complements: $\theta_2 = \frac{a_1N+b_1-b_2+\theta_1}{2}$
Strategic Independence: $\theta_2 = \frac{c-b_2}{2}$.

The strategic substitutes line is the same for both road 1 and road 2, but are sometimes disjoint and sometimes overlapping. The characterization of the best response functions are completed by finding the kink points between the three linear pieces. The kink point between the strategic substitutes and complement regions is found by their intersection. Similarly, the kink point between the strategic substitutes and independence regions is found by their intersection.

Kink Point 1: Strategic Substitutes and Complements
Road 1: $\theta_1 = (c - b_1)\frac{a_1+a_2}{2a_1+a_2}$, $\theta_2 = \frac{2a_1+2a_2}{2a_1+a_2}c - \frac{a_2}{2a_1+a_2}b_1 - a_2N - b_2$
Road 2: $\theta_1 = 2c\frac{a_1+a_2}{a_1+2a_2} - b_1 - b_2\frac{a_1}{a_1+2a_2} - a_1N$, $\theta_2 = (c - b_2)\frac{a_1+a_2}{a_1+2a_2}$

Kink Point 2: Strategic Substitutes and Independence
Road 1 Point: $\theta_1 = \frac{c-b_1}{2}$, $\theta_2 = \frac{2a_1+a_2}{2a_1}c - \frac{a_2}{2a_1}b_1 - b_2 - a_2N$
Road 1 Line: $\theta_2 = \frac{2a_1+a_2}{a_1}\theta_1 + b_1 - b_2 - a_2N$
Road 2 Point: $\theta_1 = \frac{a_1+2a_2}{2a_2}c - b_1 - a_1N - \frac{a_1}{2a_2}b_2$, $\theta_2 = \frac{c-b_2}{2}$
Road 2 Line: $\theta_2 = \frac{a_2}{a_1+2a_2}\theta_1 - \frac{a_2}{a_1+2a_2}b_2 + \frac{a_2}{a_1+2a_2}b_1 + \frac{a_1a_2}{a_1+2a_2}N$

Equilibria are characterized by one of three regimes. The first regime is duopoly when the best response functions intersect at in the region of strategic complements, when $c$ is high. The second, is the Knight-Pigou regime when the best response functions intersect in the region of strategic independence, when $c$ is low. Finally, the phantom road regime exists when the best response functions partially overlap in the strategic substitutes region, when $c$ is an intermediate value. In the first two regimes the Nash equilibrium is unique and in the third there exists a continuum of equilibria.

The lower threshold value, $\underline{c}$, is found by intersecting the lines that characterize the kink point between the strategic substitutes and independence and the strategic independence value for road 1. The higher threshold value, $\bar{c}$, is found by intersecting the kink point 1 conditions. Together these two thresholds in combination with the value of $c$ determine which of the three regimes the system is in.

Intersection Lines Characterizing Kink Point 2:
$\theta_1 = (b_2 - b_1)\frac{a_1}{a_1+a_2} + \frac{2a_2^2a_1}{(a_1+a_2)^2}N$
$\theta_1 = \frac{c-b_1}{2}$
$\underline{c} = (b_2 - b_1)\frac{2a_1}{a_1+a_2} + \frac{4a_2^2a_1}{(a_1+a_2)^2}N + b_1$

Intersection Lines Characterizing Strategic Complements
$\theta_2 = 2\theta_1 + b_1 - a_2 N - b_2$
$\theta_2 = \frac{a_1 N + b_1 - b_2 + \theta_1}{2}$
$\bar{c} = \frac{b_1(a_1+2a_2)+b_2(2a_1+a_2)}{3(a_1+a_2)} + \frac{(2a_1+a_2)(a_1+2a_2)}{3(a_1+a_2)} N$

# Appendix D The Case of Prize-Setters Competing for N Biomedical Researchers

## Appendix D.1 Overview

Suppose that there are $n$ projects to develop vaccines. Any of $N > n$ biomedical researchers may make a breakthrough if he works on one of the vaccine projects. A breakthrough on vaccine project $i$ has social value $V_i$. If $x_i$ biomedical researchers work on problem $i$, there is probability $P_i(x_i)$ that the problem will be solved, where $P_i(\cdot)$ is twice differentiable, strictly increasing, strictly concave, exogenous function bounded above by 1. Assume a biomedical researcher works on at most one problem. If no biomedical researcher works on problem $i$, it will not be solved: $P_i(0) = 0$. Assume that a planner can assign biomedical researchers to any of these $n$ projects or to the National Institutes of Health (NIH). At NIH, the social payoff per biomedical researcher is only $c$. For simplicity, assume that the Inada condition holds $P_i'(0) = \infty$. This is sufficient to insure that having *some* biomedical researchers on every project $i$ is socially optimal whether or not some researchers work at NIH. As more biomedical researchers are assigned to any given project, diminishing returns sets in. If the expected marginal social value of adding another researcher to any project is $c$ and fewer than $N$ researchers have been assigned, the remainder are assigned to NIH. If more than $N$ researchers are required to drive the expected marginal social value down to $c$, then it is optimal to assign no one to NIH.[17]

We refer to the allocation of biomedical researchers to the $n$ problems that *maximizes* the expected social payoff, $\sum_{i=1}^{n} V_i P_i(x_i) + c(N - \sum_{i=1}^{n} x_i)$ subject to $N - \sum_{i=1}^{n} x_i \geq 0$, as the socially optimal solution (denoted $x_i^s$). The first-order conditions for this Kuhn-Tucker problem are:

$$x_i \geq 0, V_i P_i(x_i) - (c + \lambda) \leq 0, \text{ with complementary slackness} \tag{D.1}$$

$$\lambda \geq 0, N - \sum_{i=1}^{n} x_i \geq 0, \text{ with complementary slackness.} \tag{D.2}$$

Given condition (D.1), the Inada assumption insures that $x_i > 0$. Hence, there are two cases. In the first, some researchers work as NIH and, since the constraint does not bind, condition (D.2) implies that $\lambda = 0$. Therefore, $V_i P_i(x_i) = c$. In the second, no researchers work at NIH and since the constraint binds, the multiplier is weakly positive. Therefore, $V_i P_i(x_i) = c + \lambda \geq c$ for $i = 1, \ldots n$. Note that in either case, it is socially optimal for the expected marginal social benefit to be equal across projects. Otherwise, a marginal expected social gain could be achieved by moving a researcher from a project with a strictly lower marginal expected social benefit to one with a strictly higher marginal expected social benefit.The maximand in the planning problem can also

[17] For additional background see Salant (2007a), Kremer and Glennerster (2004), Brennan, Macauley and Whitefoot (2011) and Salant (2007b).

be written as maximizing the social surplus on each project,

$$\sum_{i=1}^{n} x_i \left( \frac{P_i(x_i)}{x_i} V_i - c \right), \tag{D.3}$$

relative to assigning the researcher to work at NIH.

Now suppose there are $n$ independent *prize-setters*, one for each technological problem, that sets a prize $W_i$ to the first biomedical researcher to solve problem $i$. Each prize setter anticipates being able to extract the full social surplus from any discovery but must pay the researcher who made the breakthrough in project $i$, $W_i$. Hence the expected payoff of a private prize-setter is $P_i(x_i)(V_i - W_i)$. In the first stage, prize-setter $i$ sets the reward $W_i$ and in the second stage biomedical researchers decide which project to work on. Assume that each biomedical researcher working on problem $i$ attempts to solve it by himself and is as likely to solve it as anyone else working on that problem. Hence, his probability of winning the prize is $P_i(x_i)/x_i$. We denote this function as $A_i(x_i)$. Given our assumptions about $P_i(\cdot)$, each $A_i(\cdot)$ is differentiable and strictly decreasing, with $x_i A_i(x_i)$ strictly concave for $i = 1, \ldots, n$.

In the second stage equilibrium, researchers allocate themselves so that the expected payoff is the same across all projects being worked on. There are two possible conditions, one in which some biomedical researchers work for NIH,

$$W_i A_i(x_i) = c, \tag{D.4}$$

and one in which no biomedical researchers work for NIH,

$$W_i A_i(x_i) = W_j A_j(x_j). \tag{D.5}$$

The key difference with these constraints is the responsiveness of biomedical researchers to changes in a given prize. For the first constraint, when some biomedical researchers work for NIH the responsiveness of biomedical researchers to an increase in $W_i$ is increasing but depends only on the total number of researchers working on that one project:

$$\frac{dx_i}{dW_i} = \frac{-A_i(x_i)}{W_i A_i'(x_i)} > 0. \tag{D.6}$$

This follow because, as the prize-setter increases his prize, it attracts biomedical researchers away from NIH until the additional congestion on that project offsets the increased prize level; the number of researchers at the other $n - 1$ projects does not change.

In contrast, the responsiveness of biomedical researchers when no biomedical researchers work for NIH depends not only on the congestion in the prize setter's project but also on the congestion at the other projects,

$$\frac{dx_i}{dW_i} = \frac{-A_i(x_i)}{W_i A_i'(x_i) + W_j A_j'(x_j)} > 0. \tag{D.7}$$

As prize-setter $i$ increases his prize, he attracts biomedical researchers from the other projects, reducing their congestion and making them more attractive. Consequently, when no researchers work at NIH, a unit increase in the prize at project $i$ attracts fewer researchers than when they are drawn from the NIH workforce.

## Appendix D.2 Comparing the Efficient and Market Allocation of Researchers

Just as the government can set a highway-specific Pigouvian toll on each congestible road in such a way that motorists will allocate themselves to solve the planning problem so too the government can set project-specific prizes so that researchers will allocate themselves efficiently. That is, the set of socially efficient prizes is defined as the set that in the second stage induces researchers to move across projects subject to either equation (D.4) or (D.5) such that the first-order conditions for the social optimum $V_i P_i'(x_i^s) = c$ or $V_i P_i'(x_i^s) = V_j P_j'(x_j^s)$ hold, respectively. The set of socially efficient prizes, in both cases, is then given by $W_i^s = P_i'(x_i^s)V_i/A_i(x_i^s)$.

We ask whether a self-interested prize-setter would have any incentive to unilaterally deviate from the government's solution of offering the socially optimal prize $W_i^s$ in contest $i$ and attracting $x_i^s$ researchers as a result. Is the government needed as Pigou initially contended in the case of roads or, as Knight contended, is the government unnecessary in a world where self-interested independent agents set the prizes?

Consider first the case where some biomedical researchers work for NIH. In this case, the second stage constraint is given by equation (D.4) and the responsiveness of researchers to changes in the prize is given by equation (D.6).

Prize setter $i$ earns profit $\pi_i = (V_i - W_i)P_i(x_i(W_i))$ if he chooses $W_i$. Hence

$$\begin{aligned}
\partial\pi_i/\partial W_i &= -P_i(x_i) + (V_i - W_i)P_i'(x_i)\partial x_i/\partial W_i && \text{(D.8)}\\
&= \frac{1}{W_i A_i'(x_i)}\{-P_i(x_i)W_i A_i'(x_i) - A_i(x_i)V_i P_i'(x_i) + W_i A_i(x_i)P_i'(x_i)\} && \text{(D.9)}\\
&= \frac{1}{W_i A_i'(x_i)}\{-x_i A_i(x_i)W_i A_i'(x_i) - A_i(x_i)V_i P_i'(x_i)\\
&\quad + W_i A_i(x_i)[A_i(x_i) + x_i A_i'(x_i)]\} && \text{(D.10)}\\
&= \frac{A_i(x_i)}{W_i A_i'(x_i)}\{-x_i W_i A_i'(x_i) - V_i P_i'(x_i) + W_i A_i(x_i) + W_i x_i A_i'(x_i)\}, && \text{(D.11)}
\end{aligned}$$

where (D.8) is obtained by differentiating to determine the expected private gain from marginally increasing the prize, (D.9) is obtained by substituting in the researcher response in equation (D.6), (D.10) is obtained by taking account of the fact that $P_i(x) = x_i A_i(x)$, and (D.11) is obtained by factoring out $A_i(x_i)$. At $x_i^s, W_i^s = \frac{P_i'(x_i^s)V_i}{A_i(x_i^s)}$, this marginal gain from deviating from the Pigouvian prize is:

$$\frac{A_i(x_i^s)}{A_i'(x_i^s)}\left\{-V_i P_i'(x_i^s) + A_i(x_i^s)\frac{P_i'(x_i^s)V_i}{A_i(x_i^s)}\right\} = 0. \qquad \text{(D.12)}$$

Hence, whenever it is socially optimal for some researchers to work at NIH, no prize setter would have any private incentive to deviate unilaterally from the socially optimal prize in his contest provided the other prize setters adopted the socially optimal prizes in their respective contests. Hence, the profile of socially efficient prizes forms a Nash equilibrium in the first stage of the two-stage game whenever it is socially efficient for some researchers to work at NIH. This is the counterpart in contests to Knight's conclusion about congestible roads.

To understand this result, start with the profit-maximizing prize-setter's objective function $P_i(x_i)(V_i - W_i)$ and replace $W_i$ with $c/A_i(x_i)$ from the second stage constraint, given by equation (D.4). With some rearranging the profit-maximizing prize-setter's objective function can be written as $x_i(A_i(x_i)V_i - c) = V_i P_i(x_i) - cx_i$ which is the social planner's maximand in equation (D.3) and

when maximized by each independent prize-setter yields the socially efficient set of prizes.

Now consider the case where *no* biomedical researcher works for NIH. In this case, the second stage constraint is given by equation (D.5) and the responsiveness of researchers to changes in the prize is given by equation (D.7). Researchers are less responsive to an increase in the prize since as they are attracted away from the other projects, these contests become less congested. To shorten notation, denote by $Z$ the fraction $\frac{1}{W_i A_i'(x_i)+W_j A_j'(x_j)}$

$$\begin{aligned}
\partial\pi_i/\partial W_i &= -P_i(x_i) + (V_i - W_i)P_i'(x_i)\partial x_i/\partial W_i && \text{(D.13)}\\
&= Z\{-P_i(x_i)(W_i A_i'(x_i) + W_j A_j'(x_j)) - A_i(x_i)P_i'(x_i)(V_i - W_i)\} && \text{(D.14)}\\
&= A_i(x_i)Z\{-x_i W_i A_i'(x_i) - x_i A_j'(x_j)W_j - V_i P_i'(x_i) + P_i'(x_i)W_i\} && \text{(D.15)}\\
&= A_i(x_i)Z\{-W_i[x_i A_i'(x_i) - x_i A_i'(x_i) - A_i(x_i)] - x_i A_j'(x_j)W_j - V_i P_i'(x_i)\} && \text{(D.16)}\\
&= A_i(x_i)Z\{W_i A_i(x_i) - x_i A_j'(x_j)W_j - V_i P_i'(x_i)\}, && \text{(D.17)}
\end{aligned}$$

where (D.13) is obtained by differentiating to determine the expected private gain from marginally increasing the prize, (D.14) is obtained by substituting in the researcher response in equation (D.7), and (D.15) and (D.16) are obtained by taking account of the fact that $P_i(x) = x_i A_i(x)$.

Evaluated at $x_i^s, W_i^s = \frac{P_i'(x^s)_i)V_i}{A_i(x_i^s)}$, this expected marginal gain from deviating from the Pigouvian prize is:

$$A_i(x_i^s)Z\left\{\frac{-x_i^s A_j'(x_j^s)P_j'(x_j^s)V_j}{A_j(x_j^s)} - V_i P_i'(x_i^s) + \frac{A_i(x_i^s)P_i'(x_i^s)V_i}{A_i(x_i^s)}\right\} < 0, \qquad \text{(D.18)}$$

since the first factor is negative and the second factor (in braces) is strictly positive.

Hence, along the constraint where no researcher would work at NIH in the socially efficient solution, every private prize setter would have an incentive to reduce unilaterally the socially optimal prize in his contest if he conjectured that the other prize setters had adopted the socially optimal prizes in their respective contests. Hence, the profile of socially efficient prizes cannot form a Nash equilibrium in the first stage of the two-stage game in this case. Private prize setting will be inefficient.

We have found one case in which independent profit-maximizing prize-setters achieve the socially efficient allocation of biomedical researchers and another case in which they do not, depending on whether or not biomedical researchers worked at NIH.

So far, this element has been taken as exogenous. In what follows, however, we derive the best response of each prize setter as a function of the exogenous variable $c$. As we will see, there are not only the two regions we have discussed but an intermediate region where efficient equilibria are possible even though no researchers work at NIH. For simplicity we limit ourselves to the case where biomedical researchers can work at one of two projects or at NIH.

## Appendix D.3 Endogenous Determination of Whether Researchers Work at NIH

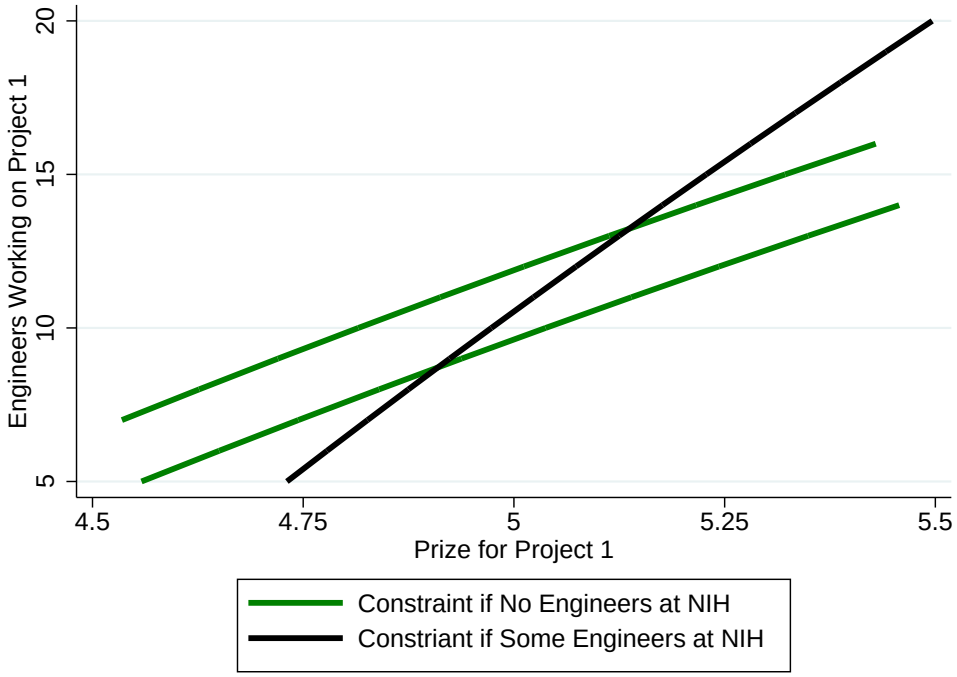
First, consider that the prize for project 1, given a conjectured prize for project 2, induces biomedical researchers to work at both projects and NIH. If the prize-setter for project 1 marginally increases the prize biomedical researchers will switch to working on project 1 until the expected benefit of working on project 1 decreased to equal the benefit of working at NIH. All of the additional

biomedical researchers working on project 1 previously worked at NIH, since the conjectured prize for project 2 remained unchanged and thus the probability the project is solved must also remain unchanged for the expected benefit to remain equal to the benefit of working at NIH. The case described here is exactly the case governed by equations (D.4) and (D.6) discussed above.

Eventually, however, as the prize for project 1 increases there will be no biomedical researchers left working at NIH. At that point, raising the prize on project 1 must cause researchers to switch from project 2. On the interior of this region the two projects have a common expected benefit for the biomedical researchers which is strictly higher than the benefit of working at NIH. This region is governed by equations (D.5) and (D.7) where researchers are less responsive to changes in the prize on project 1 than they were in the region governed by equations (D.4) and (D.6). Therefore the constraint is kinked at the boundary between these two regions.

Figure D.1 graphs the constraint for when some biomedical researchers work at NIH (black solid line) and two constraints when no biomedical researchers work at NIH for two different levels of the prize for project 2 (green solid line). The further southeast of the two constraints when no biomedical researchers work at NIH corresponds to a larger prize for project 2. The constraint when some biomedical researchers work at NIH is steeper than the constraints if no biomedical researchers work at NIH because the biomedical researchers are more responsive if the change in researchers to project 1 is due to changes in researchers at NIH where the congestion on the other projects is unaffected, equations (D.6) and (D.7).

Figure D.1: Two Projects and NIH Constraints




In the first stage the prize-setter maximizes their expected payoff, $P_i(x_i)(V_i - W_i)$. With some

probability their project is solved in which case they receive the social benefit $V_i$ and must pay the biomedical researcher that solved the problem the prize $W_i$. Prize-setters balance the benefits of increasing the probability the problem is solved with the cost of increasing the prize. The slope of the indifference curve for the prize-setter is given by differentiating the prize-setter's objective function $U = P_i(x_i)(V_i - W_i)$ which is,

$$\frac{dx_i}{dW_i} = \frac{P_i(x_i)}{P_i'(x_i)(V_i - W_i)} > 0. \tag{D.19}$$

Since points to the northeast have both higher $x_i$ and higher $W_i$, the numerator is larger, the denominator is smaller, and hence the quotient is larger. Each indifference curve is therefore positively sloped and strictly convex.

Figure D.2: Two Projects, NIH Constraints, and an Indifference Curve

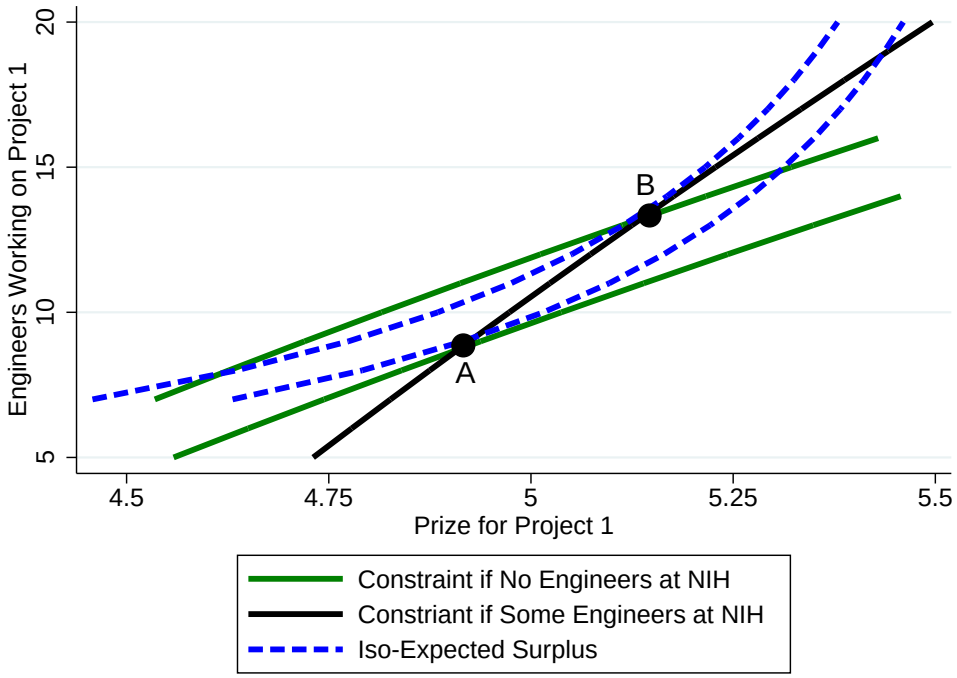



Figure D.2 graphs the two constraints from the second stage and the indifference curve from the first stage. Point A represents the point at which the constraints intersect and the indifference curve is tangent to the constraint when all researchers work on congestible projects. Point B represents a different point where the constraints intersect (a lower prize for project 2) and the indifference curve is tangent to the constraint when some researchers work at NIH. Three cases exist; one with a low prize for project 2 where the constraints intersect northeast of point B, one with an intermediate prize for project 2 where the constraints intersect between points A and B, and one with a high prize for project 2 where the constraints intersect southwest of point A.

As the prize for project 2 decreases in the first case, where the prize for project 2 is low, the

tangency of the indifference curve remains at point B. The best prize for the profit-maximizing prize-setter on project 1 to set remains constant for all prize levels for project 2 lower than the prize level with the constraint that intersects point B. In this region the best response function for the profit-maximizing prize-setter on project 1 is strategically-independent of prize 2.

In the second case, where the prize for project 2 is an intermediate level, the indifference curve intersects the constraint at the kink between the constraints. As the prize for project 2 decreases the kink point moves northeast implying the best reply for the profit-maximizing prize-setter on project 1 is to increase their prize. In this region the profit-maximizing prize-setter on project 1's prize is a strategic-substitute of the prize for project 2.

Finally in the third case, where the prize for project 2 is high such that no researchers work for NIH the indifference curve is tangent to the flatter constraint given by equation (D.5). In this region the prizes can be strategic-substitutes or strategic complements.

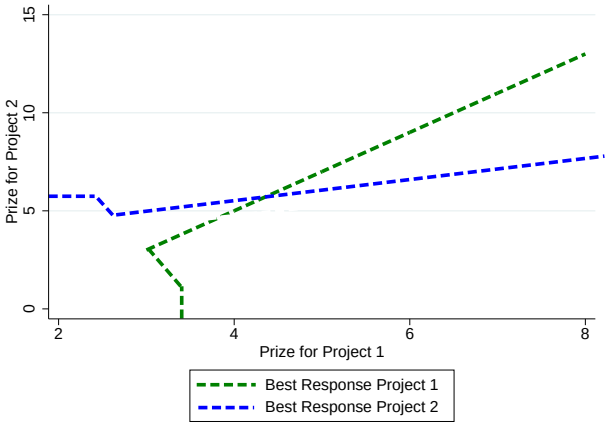
## Appendix D.4 Best Response Functions

This section graphs the best response functions of the prize-setters. The previous subsection demonstrated that the best response functions exhibit regions of strategic-independence, strategic-substitutes, and can exhibit regions of strategic-complements. Figure D.3 graphs six cases for the best response functions. The figures on the left graph the best response functions when there is a region that exhibits strategic-complements, while the figures on the right graph the best response functions when the best response functions exhibit only strategic-independent and strategic-substitutes regions. The figures on the top, middle, and bottom graph the best response functions when the wage at NIH is low, intermediate, and high, respectively.

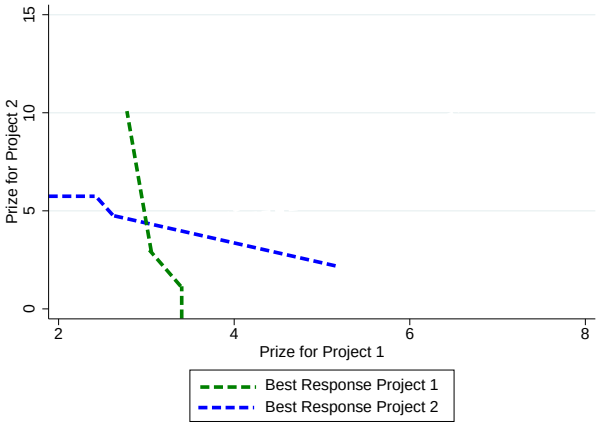
When the wage at NIH is low, depicted in the top two Figures of Figure D.3, the best response functions intersect in the region of strategic-complements (left graph) or strategic-substitutes (right graph). In both cases all researchers work on projects 1 and 2 and no researchers work at NIH. When the wage at NIH is high, depicted in the bottom two Figures of Figure D.3, the best response functions intersect in the region of strategic-independence. In this case researchers work at both projects and at NIH.

When the wage at NIH is an intermediate level, depicted in the middle two Figures of Figure D.3, the best response functions overlap in a region characterized by strategic-substitutes. In this case all of the researchers work on projects 1 and 2 but the wage at NIH disciplines the market because if either prize-setter unilaterally lowered their prize some researchers would switch and work for NIH. In this case there are a continuum of equilibria which may or may not include prizes that induce the efficient allocation of researchers across projects.

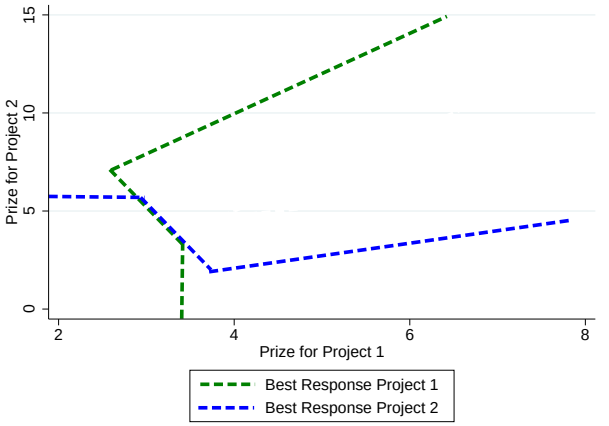
Figure D.3: Best Response Functions



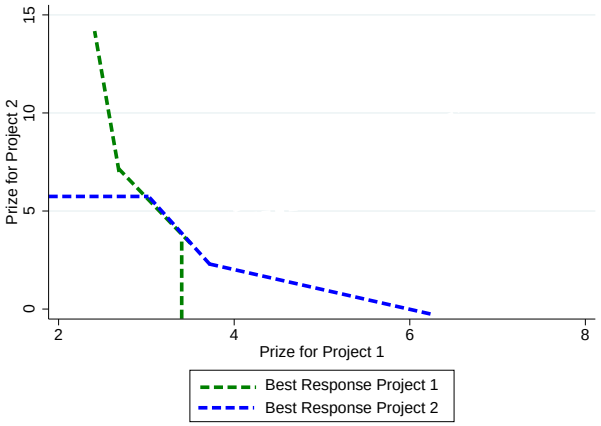

Low Wage NIH: Strategic Complements



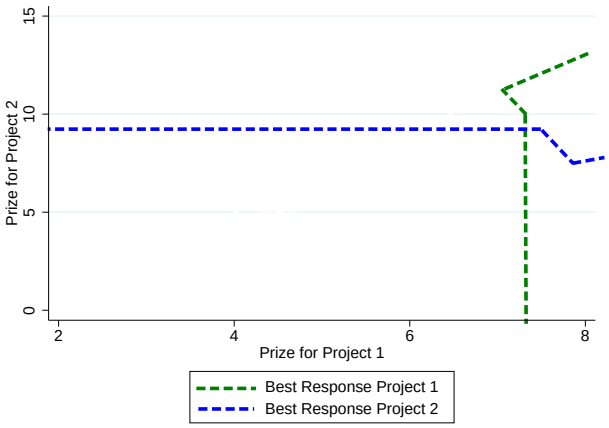

Low Wage NIH: Strategic Substitutes



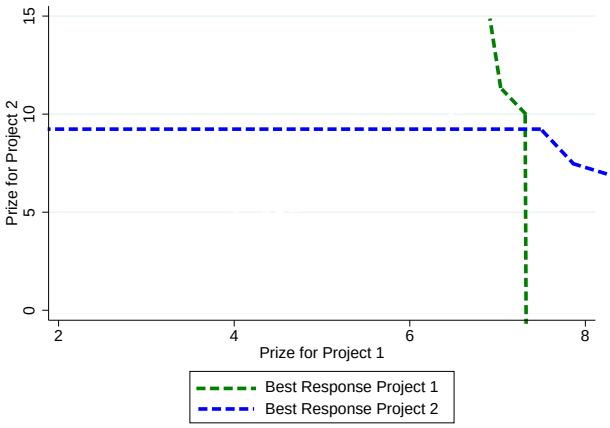

Intermediate Wage NIH: Strategic Complements




Intermediate Wage NIH: Strategic Substitutes




High Wage NIH: Strategic Complements




High Wage NIH: Strategic Substitutes